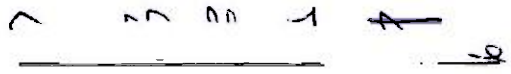
In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.



7/25/68

STATISTICAL TOLERANCE LIMITS FOR A PEARSON TYPE III DISTRIBUTION

A THESIS

Presented to

The Faculty of the Graduate Division

by

Darrell Glenn Fontane

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Civil Engineering

Georgia Institute of Technology

March, 1970

STATISTICAL TOLERANCE LIMITS FOR A PEARSON TYPE III DISTRIBUTION

Approved:

Chairman

Date approved by Chairman: March 18, 1970

ACKNOWLEDGMENTS

The author wishes to express sincere gratitude to Dr. J. R. Wallace, thesis advisor. Throughout the development of this thesis, Dr. Wallace devoted much time and effort in assisting the author.

Sincere appreciation is also given to Dr. W. W. Hines and Dr. G. M. Slaughter. The comments and suggestions they provided were very helpful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

GLOSSARY OF TERMS

Terms

u	is the population mean.
σ^2	is the population variance.
σ	is the population standard deviation (equal to the square root of the variance).
L	is the population limit.
G	is the population skew.
P90	is the population 90 per cent limit.
$\bar{x}$	is the sample mean (equation (23), page 26).
s^2	is the sample variance (equation (25), page 26).
s	is the sample standard deviation (equal to the square root of the sample variance).
g	is the sample skew (equation (26), page 27).
S90	is the sample 90 per cent limit (equation (27), page 28).
n	is the sample size.
X^2	denotes the chi-square distribution.

SUMMARY

In 1967 the Water Resources Council recommended that federal agencies adopt the log-Pearson Type III distribution as the base method for flood flow frequency analysis. Flood frequency analysis is a statistical prediction of future events. From a sample of flood data, specified recurrence interval floods can be estimated. Since these specified recurrence interval floods are determined from a sample of all possible floods, they can be expected to vary in the future. Statistical tolerance limits provide a means of estimating the range of future variation in a specified recurrence interval flood. The purpose of this thesis was to develop statistical tolerance limits for the Pearson Type III distribution.

Development of tolerance limits depends upon the determination of tolerance factors. Determination of tolerance factors in turn depends upon knowing the distribution of the ratio of sample to population variance. An analytical approach to determine the distribution of the ratio of variances failed because of the difficulty of the mathematics involved. A numerical technique was then tried. By simulation on a digital computer, samples from a Pearson Type III distribution were generated. A set of empirical tolerance factors was developed directly from the generated data. These factors

are intended to serve as guideline estimates for future work.

CHAPTER I

INTRODUCTION

In December of 1967 the Water Resources Council published Bulletin No. 15 entitled "A Uniform Technique for Determining Flood Flow Frequencies". The purpose of this bulletin was to propose the adoption by federal agencies of the log-Pearson Type III distribution as the base method for flow frequency analysis. The objective of this thesis is the determination of statistical tolerance limits for the Pearson Type III distribution.

Flow Frequency Analysis

Flow frequency analysis is a method of statistical prediction of future events. A record of flood flows is used to develop a flood frequency curve. From this frequency curve, the average number of future years which will experience floods equal to or greater than a specified flood event can be designated (Linsley et al., 1958, pp. 248-250)*. For example, a 100-year flood is that flood which, on the average, will be experienced in one per cent of all future years.

A designated flood such as the 100-year flood is actually a percentile of the sample. One method used to

---

*Refers to reference listed in Appendix E.

determine the average number of years which will experience a particular magnitude of flooding (Snyder, 1966, pp. 20-21) is given by the expression

$$PE = \frac{m}{n + 1} \qquad (1)$$

where PE is the average probability of occurrence
n is the sample size or length of record
and m is the rank of the flood where all floods are ranked in descending order of magnitude.

Therefore, the largest flood in a sample of nine floods would have an estimated average probability of occurrence of ten per cent which corresponds to the 10-year flood. Thus, it is estimated that, on the average, 90 per cent of all future floods will be smaller than the 10-year flood.

The computed percentiles or proportions can be expected to vary with future samples. A record of flood events is only a sample of all floods, past and future. Therefore, a frequency curve developed from that record is also a sample. As the flood record changes with time, the frequency curve will also change. It is desirable to estimate the variation of future frequency curves and hence the variation of the computed percentiles and proportions. Statistical tolerance limits provide probabilistic ranges for this variation.

Statistical tolerance limits are those limits between, above, or below which one expects to find a specified proportion of the population, a specified per cent of the time (Natrella, 1963, p. 2-13). In computation of tolerance limits for flood flows, a limit is set such that a specified proportion of the sample, P, will be smaller than that limit in β per cent of future flood samples (Snyder, 1966, p. 25). This tolerance limit is computed by

$$Q_u = \overline{Q} + Ks \qquad (2)$$

where $\overline{Q}$ is the sample mean flood (calculated from the sample data)

s is the sample standard deviations of floods (calculated from the sample data)

and K is a factor dependent on the proportion, P, the probability, β, the distribution of the random variable Q, and the sample size, n.

For a given sample and fixed values of K, values of Q_u can be computed for various values of P and β. These values of Q_u form limit curves to the flood frequency curve. A flood frequency curve and a set of tolerance limit curves have been computed for a hypothetical flood record. These curves are plotted in Figure 1. The sample of flood flows was assumed to be normally distributed for which values of K are available (Natrella, 1963, pp. T-14 - T-15).

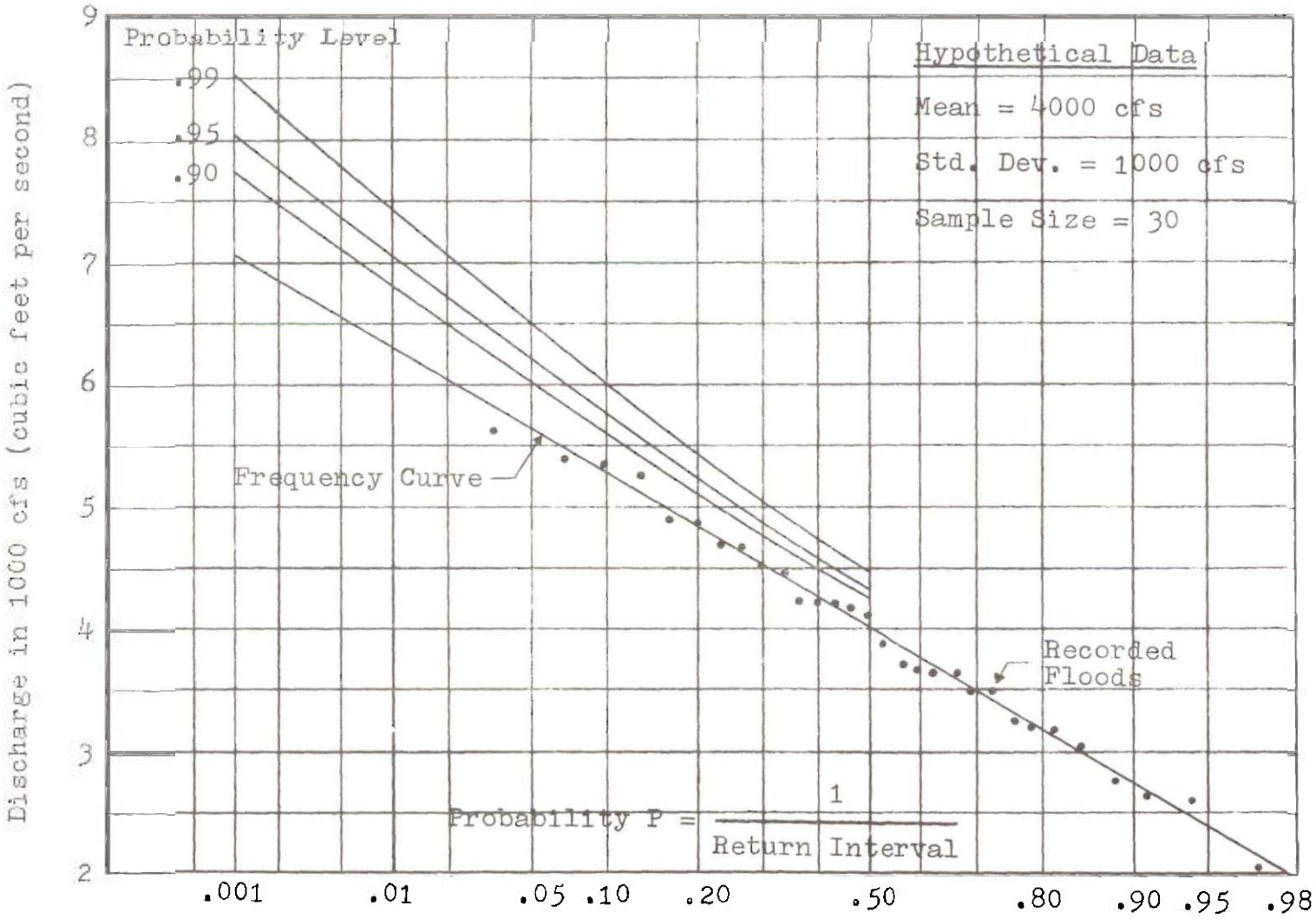


Figure 1. Tolerance Limits of a Frequency Curve.

Figure 1 indicates that the flood frequency curve is not necessarily the flood frequency curve for all floods. The curve of future floods may be noticeably different from the one shown. However, that future curve is expected to lie below the limit curves with probability as noted.

The magnitude of a specified recurrence interval flood is not necessarily a constant value. There is some chance that a flood of given recurrence interval could have a larger magnitude than indicated by the flood frequency curve. This chance is measured by the probability levels on the tolerance limit curves. From the frequency curve (Figure 1), the 100-year flood (plotted at the probability of occurrence equal to one per cent) has a magnitude of 6300 cfs, where cfs denotes cubic feet per second. However, from the limit curves, the 100-year flood has a one per cent chance of being larger than 7446 cfs.

In the same way that variability exists for the magnitude of a given recurrence interval flood, the recurrence interval of a given flood is also variable. Again consider the 100-year flood from the frequency curve. From Figure 1, a flood of that magnitude (6300 cfs) has a five per cent chance of having a probability of occurrence of four per cent. In other words, the 100-year flood from the frequency curve has a five per cent chance of having a recurrence interval of 25 years.

The purpose of the preceding discussion has been to

illustrate that based on a frequency curve alone it is incorrect to state that:

"The Nth-year flood has a magnitude of X cfs."

It is more appropriate to state that:

"The Nth-year flood has probability β of being larger than X cfs."

The use of a probabilistic range, i.e., tolerance limits, in analysis of flood flow frequency adds flexibility to planning and design. Safety depends upon minimizing the risk of floods higher than the design flood. A design recurrence interval flood can be selected from this probabilistic range of values in accordance with the allowable risk appropriate to the purpose of the design.

The Pearson Type III Distribution

Numerous statistical distributions have been used in flow frequency analysis (Bulletin 13, 1966). Bulletin No. 15 (1967) was the result of an effort to standardize flow frequency analysis by federal agencies. Because of its flexibility in application the log-Pearson Type III distribution was recommended as the base method for flow frequency analysis.

One of the earliest applications of the Pearson Type III distribution to hydrologic problems was presented in 1924 by H. A. Foster. Foster reasoned that a frequency curve which was limited in one direction and skewed should be used for

stream-flow studies. The choice of this type of curve was based on the fact that runoff can not have a value less than zero and also has no definite upper bound. The Pearson Type III curve satisfied Foster's requirements.

The Pearson Type III curve, developed by Karl Pearson, is one of a series of probability functions and has the mathematical form (Elderton et al., 1969, pp. 78-81)

$$y = y_o(1 + x/a)^p e^{-px/a} \qquad -a \leq x < \infty \qquad (3)$$

with $p = \gamma a$

$$y_o = \frac{p^{p+1}}{ae^p \Gamma(p+1)}$$

where p is the skewness parameter or that parameter indicating the degree of departure from symmetry

a is the lower bound of the curve

y_o is the value of the ordinate of the curve at the mode, that is, at $x = 0$

and $\Gamma(p+1)$ is the complete gamma function (See Appendix B).

The origin of the axis is at the mode and p, γ, and a are the three distribution parameters. A detailed discussion of this distribution in regard to its form, its parameters, and some of its characteristics is presented in Appendix A. The

Pearson Type III distribution has the general shape* shown in Figure 2.

The log-Pearson Type III distribution recommended in Bulletin No. 15 (1967) has the same mathematical form as the Pearson Type III distribution. The nomenclature of log-Pearson arises from the technique used to fit a Pearson Type III curve to data. In Foster's work, untransformed data were used to fit the curve. The recommendations of Bulletin No. 15 (1967, p. 7) require a logarithmic transformation; the curve is then fitted to the logarithms. Because of this transformation, the resulting curve is called the log-Pearson Type III.

Statistical Tolerance Limits

There are two types of statistical tolerance limits; those which depend on the underlying distribution and those which do not (Natrella, 1963, pp. 2-13 - 2-15). The latter are called distribution-free tolerance limits. Distribution-free limits are wider for a given sample size than those limits based on the underlying distribution (Natrella, 1963, p. 2-15). Also, distribution-free limits require large sample sizes relative to distribution-dependent limits for reasonable probability statements to be made (Bowker et al., 1959, pp. 229-232).

---

*The curve is usually bell-shaped as shown in Figure 2, but becomes J-shaped when the skewness exceeds a value of one.

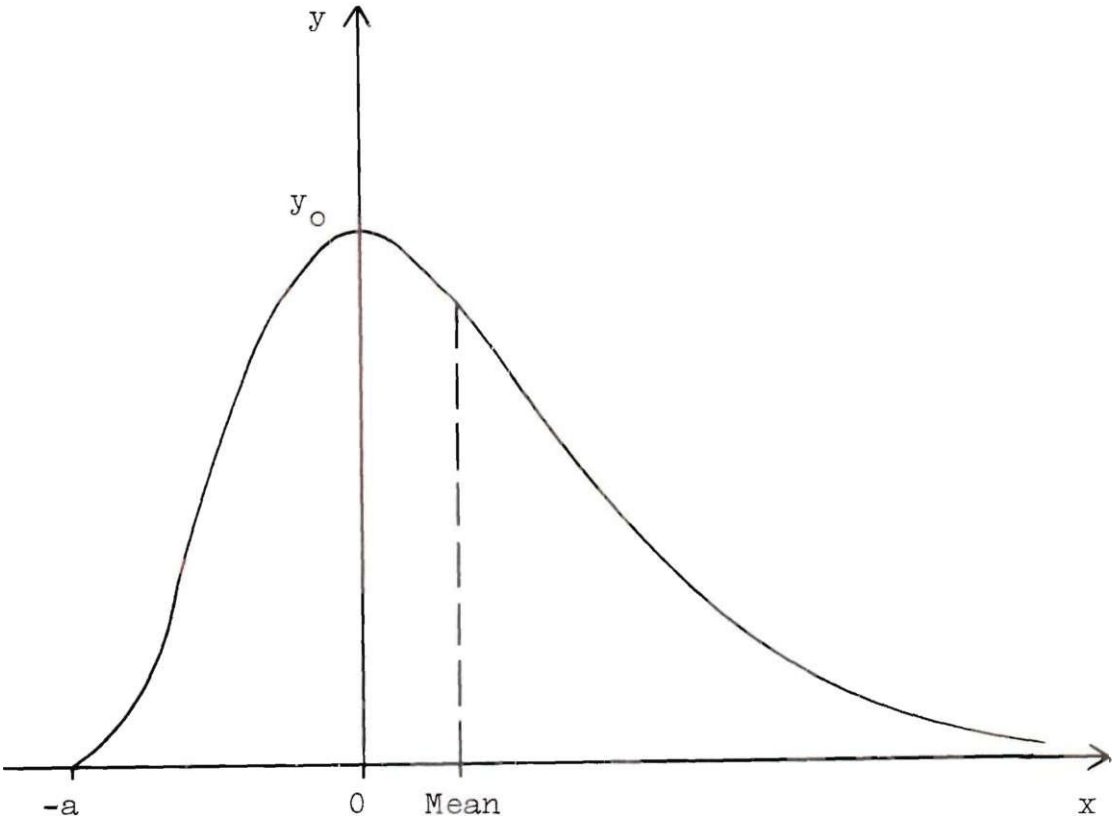


Figure 2. The General Form of a Pearson Type III Curve for Skewness Less Than One.

Statistical tolerance limits have been developed for the normal distribution. There are two catagories of statistical tolerance limits. These are two-sided tolerance limits and one-sided tolerance limits. Two-sided normal tolerance limits were developed by Wald and Wolfowitz in 1946. Two-sided tolerance limits are those limits such that with probability, β, at least a proportion of the population, A, lies within the limits. Two-sided limits are computed by

$$LL = \bar{x} - \lambda s \quad \text{(lower limit)} \tag{4}$$

$$UL = \bar{x} + \lambda s \quad \text{(upper limit)} \tag{5}$$

where $\bar{x}$ is the sample mean

s is the sample standard deviation

and λ is a constant termed the tolerance factor.

Figure 3 is a graphical representation of the two-sided tolerance limit. The area between the lower and upper limits, i.e., between $\bar{x} \pm \lambda s$, is equal to "a". The area between the population limits PL and PU is equal to "A". The probability that the shaded area, a, is greater than the specified population area, A, is equal to β or

$$P(a \geq A) = \beta$$

Because the normal distribution is symmetrical about

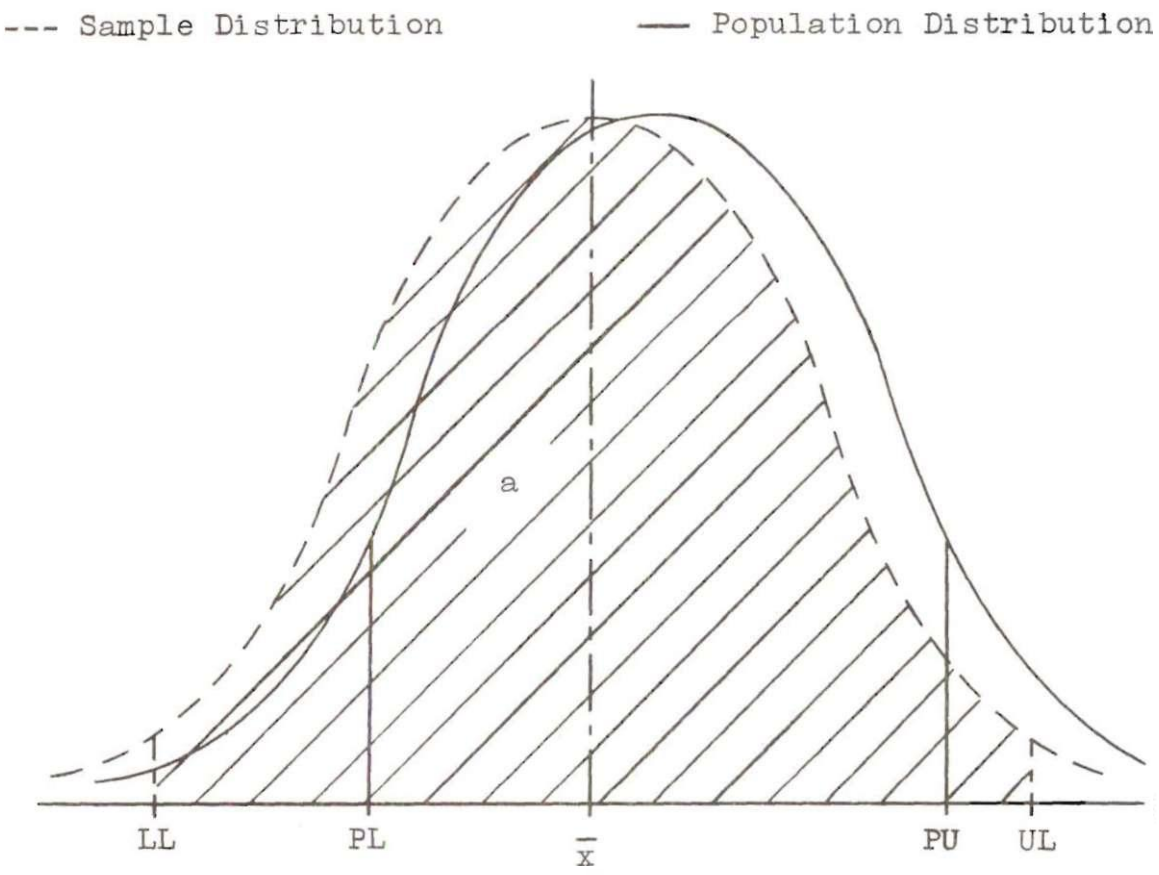


Figure 3. A Two-Sided Tolerance Limit.

the mean, the lower and upper limits can be specified in terms of a single tolerance factor, λ. For a non-symmetric distribution, such as the Pearson Type III, two sets of factors, one for lower limits and the other for upper limits, would have to be used.

One-sided statistical tolerance limits for a normal distribution were developed by Johnson and Welch in 1940. A one-sided upper tolerance limit is that limit, l_u, such that the probability that at least a proportion, P, of the population is less than l_u, is equal to epsilon, E. For a given population there is a limit, L, such that a proportion, P, of the population is less than that limit. A one-sided upper tolerance limit can then be redefined as that limit, l_u, such that with probability, E, the limit, l_u, is greater than or equal to the population limit, L.

In mathematical notation

$$P(l_u \geq L) = E \tag{6}$$

The tolerance limit, l_u, is computed by

$$l_u = \bar{x} + Ks \tag{7}$$

where $\bar{x}$ is the sample mean

s is the sample standard deviation

and K is a constant termed the one-sided tolerance factor.

A one-sided upper tolerance limit is represented by Figure 4.

Correspondingly a one-sided lower tolerance limit is that limit, l_1, such that the probability that at least a proportion, P, of the population is greater than l_1 is equal to E. The lower one-sided tolerance limit, l_1, is computed by

$$l_1 = \bar{x} - Ks$$

As in the case of the two-sided tolerance limit, the symmetry of the normal distribution about its mean allows for the use of a single constant, K, for both the upper and lower one-sided limits. A non-symmetric distribution would require two sets of constants.

Orientation of Research

In the previous discussion of flow frequency analysis, the use of tolerance limits for flood flow analysis was illustrated. These tolerance limits were computed by equation (2). Tolerance limits so computed are one-sided upper tolerance limits. Because flow frequency analysis and in particular flood flow analysis was the intended area of application of the results of this thesis, the research was oriented towards the one-sided upper tolerance limit. Also, this type of tolerance limit requires the evaluation of only one set of

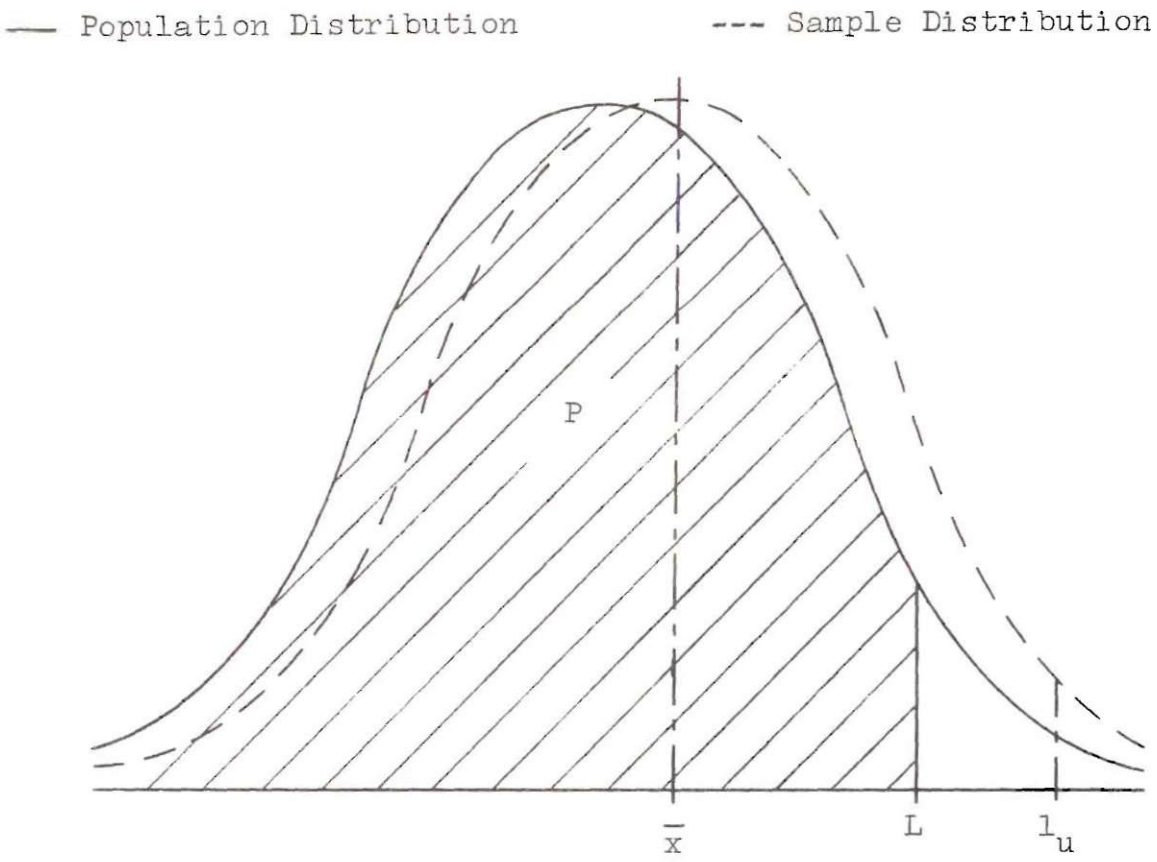


Figure 4. A One-Sided Upper Tolerance Limit.

tolerance factors.

Evaluation of a tolerance limit depends upon the evaluation of the tolerance factor. From equation (7) it can be noted that $\bar{x}$ and s are computed from the sample and the constant K is the only remaining term required. Because of this, the constant K must be a function of the degree of probability statement, E, the population limit, L, which is a measure of the required proportion, and the sample size used to compute $\bar{x}$ and s. In their development of the tolerance factors for a one-sided tolerance limit, Johnson and Welch (1940) made use of the non-central "t" statistic. The non-central "t" statistic is defined by the equation (Johnson et al., 1940, p. 362)

$$t = \frac{z + \delta}{w^{\frac{1}{2}}} \tag{8}$$

where z is a quantity distributed normally about zero with unit standard deviation

w is a quantity distributed independently as X^2/f, where X^2 represents the chi-square distribution and f is the number of degrees of freedom

and δ is a constant.

This statistic is distributed in a manner that depends only on δ and f.

The computational form for a one-sided upper tolerance limit is given by equation (7)

$$l_u = \bar{x} + Ks$$

A similar expression can be written in terms of the population limit, L

$$L = \bar{x} + k_u s \tag{9}$$

There is a particular value of k_u that will satisfy equation (9) for any given sample. Rewriting equation (9)

$$k_u = \frac{L - \bar{x}}{s} \tag{10}$$

It should be noted that $\bar{x}$, s, and k_u are random variables and L is a constant related to the proportion, P. Multiplying equation (10) by $n^{\frac{1}{2}}$ (n is the sample size) yields

$$n^{\frac{1}{2}} k_u = n^{\frac{1}{2}} \frac{(L - \bar{x})}{s} \tag{11}$$

Further manipulation of equation (11) gives

$$n^{\frac{1}{2}}k_u = \left(\frac{n^{\frac{1}{2}}}{\sigma}(L-u) - \frac{n^{\frac{1}{2}}}{\sigma}(\bar{x}-u) \right) \div \frac{s}{\sigma} \qquad (12)$$

where u is the population mean

and σ is the population standard deviation.

Equation (12) is equivalent to equation (11). For any distribution, the population limit, L, mean, u, and standard deviation, σ, are constants. Therefore, $\frac{n^{\frac{1}{2}}}{\sigma}(L-u)$ is a constant term and is equivalent to δ in equation (8). The sample means from any distribution are approximately normally distributed with a mean equal to the population mean, u, and a standard deviation equal to the population standard deviation divided by the square root of the sample size, $\sigma/n^{\frac{1}{2}}$ (Burington et al., 1958, p. 151). Therefore, the term $\frac{n^{\frac{1}{2}}}{\sigma}(\bar{x}-u)$ will be normally distributed with a mean equal to zero and a standard deviation equal to one. This is equivalent to z in equation (8). For a normal distribution, a function of the ratio of sample to population variances is chi-square (X^2) distributed. More precisely, $(n-1)s^2/\sigma^2$ is distributed as X^2 with (n-1) degrees of freedom (Guttman et al., 1965, p. 144). Therefore, for the normal distribution, s^2/σ^2 is distributed as X^2/f, which is equivalent to w in equation (8). For the normal distribution then, equation (12) is equivalent to equation (8) and $n^{\frac{1}{2}}k_u$ has a non-central "t" distribution

$$n^{\frac{1}{2}}k_u = t(n-1,\ n^{\frac{1}{2}}U,\ E) \qquad (13)$$

where $(n-1)$ is the number of degrees of freedom

$n^{\frac{1}{2}}U$ is a measure of the required proportion or limit as $U = (L-u)/\sigma$

and E is the probability or confidence desired.

Given a particular sample size, proportion or limit, and probability, the corresponding tolerance factor can be computed from

$$K = \frac{t(n-1,\ n^{\frac{1}{2}}U,\ E)}{n^{\frac{1}{2}}} \qquad (14)$$

Tolerance factors of this type have been tabulated in terms of these three parameters (Natrella, 1963, pp. T-14 - T-15).

It is important to note that in relating equation (12) to equation (8) the equivalence of the terms z and δ did not depend on the population distribution. However, the equivalence to the w term could only be made because the distribution of a function of the sample variance for a normal distribution was known.

In order to develop tolerance limits for a Pearson Type III distribution in a manner analagous to that used by Johnson and Welch (1940) for the normal distribution, the distribution of the sample variance of the Pearson Type III

must be known. The first phase of the research was directed toward an analytical determination of this distribution. This phase of the research was not successful. Therefore, a different approach was taken in subsequent phases of the research. Specifically, simulation techniques employing a digital computer were used to generate samples from Pearson Type III distributions. These samples provided the empirical data required to establish the distribution of various statistics of the Pearson Type III distribution. Ultimately, the data were used to experimentally determine statistical tolerance limits.

CHAPTER II

PROCEDURE

The research was carried on in three phases. The objective of the first two phases was to determine one-sided upper statistical tolerance limits for the Pearson Type III distribution in a manner analagous to that used by Johnson and Welch (1940) for the normal distribution. The objective of the third phase of the research was to estimate the tolerance factors for a Pearson Type III distribution by determining a set of empirical tolerance factors based on generated samples from a Pearson Type III distribution.

Research Phase I

The first phase of the research was an analytical approach to the determination of the distribution of the sample variance from a Pearson Type III distribution. In order to pursue this approach, the density function of the sample variance must be known. The density function of the sample variance, $f(s^2)$, can be specified in terms of a cumulative probability statement. This can be expressed as follows.

$$P\left(\frac{\Sigma(x_i - \bar{x})^2}{n - 1} \leq s^2\right) = F(s^2) \qquad (15)$$

where $F(s^2)$ is the cumulative probability function and the x_i's are independent and distributed as the Pearson Type III distribution.

Rearranging terms in equation (15) gives

$$P\left(\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] \leq (n-1)s^2\right) = F(s^2) \qquad (16)$$

The density function is equal to the derivative of the cumulative function

$$f(s^2) = F'(s^2) \qquad (17)$$

The density function of the sample variance for a normal population can be found by the use of the Moment Generating Function (Finney et al., 1968, pp. 38-40, 99-103). An analagous approach was tried for the sample variance from a Pearson Type III population. Also tried was the determination of the sample variance distribution by consideration of the problem as a function of random variables. Both of these approaches failed because of the difficulty of the mathematics involved. The solution of equation (17), where $F(s^2)$ is defined by equation (16) and the x_i's are distributed as a Pearson Type III distribution, will be very difficult. Rather than trying to determine mathematical

approximations that could be used for the solution of equation (17), the research was directed toward numerically evaluating the distribution of the sample variance.

Research Phase II

The second phase of the research involved simulation as a means of determining the sample variance distribution. Although there are simulation or "Monte Carlo" techniques which can be used to generate samples from a Pearson Type III population, for the purpose of this work the Pearson Type III distribution was transformed to an equivalent distribution. By a change of variable, the form of the Pearson Type III distribution reduces to that of the Gamma distribution. The proof of this transformation is given in Appendix B. The Gamma distribution has the form (Naylor et al., 1966, pp. 87-89)

$$y = \frac{\alpha^k x^{(k-1)} e^{-\alpha x}}{\Gamma(k)} \qquad 0 \leq x < \infty \qquad (18)$$

where α is the scale parameter

k is the shape or skewness parameter

and $\Gamma(k)$ is the complete gamma function which is equal to (k-1) factorial for integer values of k.

For this distribution, the mean, u, is equal to k/α, the

variance, σ^2, is equal to k/α^2, and the skew, G, is equal to $1/k^{\frac{1}{2}}$.

Computer Simulation

The use of the Gamma distribution has the advantages of a simpler mathematical form and the availability of standardized techniques for generation. The technique used to generate samples from a Gamma distribution was the one given in Computer Simulation Techniques (1966, pp. 87-89) by Naylor, Balintfy, Burdick, and Chu. The probability distribution of the sum of k independent exponential variates each with parameter α will be a gamma distribution with parameters α and k. This combination yields the mathematical expression

$$x = -\frac{1}{\alpha}\left(\log \prod_{i=1}^{k} r_i\right) \qquad (19)$$

where x is distributed according to a Gamma distribution with parameters α and k

r_i is a random number uniform on the interval from zero to one

and π refers to the product of terms.

For convenience, the scale parameter, α, was set equal to one. Also, as shown in Appendix B, the one-sided tolerance factor is independent of α.

Equation (19) provides a method for generating variates and hence samples from a Gamma distribution. Generated

samples were used to determine the distribution of sample statistical parameters. From a generated sample, selected statistical parameters can be computed. A second sample can then be generated and a second set of values of the parameters determined. If this procedure is repeated, a number, or set, of values for each computed parameter will be obtained. The distribution of these values can then be determined, for example, by grouping the data and plotting its histogram.

Preliminary to the generation of data, two items had to be determined: First, the statistical parameters whose distributions would be evaluated, and second, the computational forms used to compute those parameters.

Statistical Parameters Selected

Although the objective of the second phase of the research was the determination of the sample variance distribution, the distributions of the sample means, sample skews, and sample 90 per cent limits were also determined. The distribution of the sample means was found as a "check" on the theoretical distribution of sample means (see Chapter I, page 17). A tolerance limit for a Pearson Type III distribution will be a function, in some manner, of the skew of that distribution. Therefore, the distribution of the sample skews was determined. Finally, a one-sided tolerance limit is an estimate of the variability of a sample limit. An indication of this variability will be given by the distribution of the

sample limit. The distribution of the sample 90 per cent limit was arbitrarily chosen for study.

Parameter Computation

For the normal distribution the maximum likelihood estimates of the mean and variance are (Markovic, 1965, p. 8)

$$\text{Mean} = \bar{x} = (1/n) \sum_{i=1}^{n} x_i$$

$$\text{Variance} = s^2 = (1/n) \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Johnson and Welch (1940) in their development of one-sided tolerance limits for the normal distribution used the unbaised estimate for the sample variance

$$s^2 = (1/(n-1)) \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (20)$$

For the Gamma distribution, the estimates of the sample mean and variance are functions of the estimates of the parameters α and k. The maximum likelihood estimates of α and k are (Markovic, 1965, pp. 8-9)

$$\hat{\alpha} = \hat{k}/(\frac{1}{n} \sum_{i=1}^{n} x_i) \qquad (21)$$

$$\hat{k} = \frac{1 + \left(1 + \frac{4}{3}\left(\ln \bar{x} - \frac{1}{n}\sum_{i=1}^{n} \ln x_i\right)\right)^{\frac{1}{2}}}{4\left(\ln \bar{x} - \frac{1}{n}\sum_{i=1}^{n} \ln x_i\right)} - \Delta\hat{k} \qquad (22)$$

where $\bar{x} = (1/n) \sum_{i=1}^{n} x_i$

and $\Delta\hat{k}$ is a tabulated correction factor.

The estimates of the sample mean and variance can now be defined by

$$\text{Mean} = \bar{x} = \hat{k}/\hat{\alpha} = (1/n) \sum_{i=1}^{n} x_i \qquad (23)$$

$$\text{Variance} = s^2 = \hat{k}/\hat{\alpha}^2 = \bar{x}^2/\hat{k} \qquad (24)$$

Since the computational form for $\hat{k}$ given in equation (22) must be used in equation (24) to compute the sample variance, this makes the estimate of the sample variance computationally difficult. Since the maximum likelihood estimate of the sample variance is computationally unsatisfactory, the estimate of the sample variance recommended in Bulletin No. 15 (1967, p. 8) was used. This estimate is

$$s^2 = (1/(n-1)) \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (25)$$

Equation (25) is identical to equation (20), the estimate of Johnson and Welch (1940). This work was oriented toward flow frequency analysis, therefore the estimate of the sample variance recommended in Bulletin No. 15 (1967) for use in fitting a Pearson Type III curve to data appears to be the most logical estimate to use.

To obtain an indication of the difference in the estimates of the sample variance (equations (24) and (25)), data from the work of Markovic (1965) were analyzed. Markovic (1965, pp. 30-33) determined the maximum likelihood estimates for the Gamma parameters α and k, from a sample of river flow data. Using these estimates and data, the difference in the sample variance computed by equations (24) and (25) was found to be less than 2.2 per cent.

The sample skew was computed by first computing the coefficient of skewness by the method recommended in Bulletin No. 15 (1967, p. 8). Foster (1924) has shown that the skew is equal to one-half the coefficient of skewness (Foster, 1924, p. 154) and this yields the equation for the skew

$$g = (1/2)\frac{n \sum_{i=1}^{n} (x_i - \bar{x})^3}{(n-1)(n-2)s^3} \qquad (26)$$

where s is the sample standard deviation computed from equation (25); the standard deviation is

the square root of the variance.

The sample 90 per cent limit was computed by

$$\text{Per cent limit} = \frac{m}{n + 1} \tag{27}$$

where m is the rank of a data point where the data are ranked in ascending order of magnitude and n is the sample size.

Random Number Generation

The computer used for the simulation was the UNIVAC 1108. Available on the UNIVAC 1108 system were two groups of subroutines, the Math-Pack and the Stat-Pack, portions of which were used for this work. All programming was done in the Fortran IV language.

The algorithm used to generate gamma variates, equation (19), requires random numbers uniform on the interval from zero to one, U(0,1), as input. The random numbers were obtained from the RANDU subroutine in the Math-Pack. This subroutine produces random numbers U(0,1). The randomness and uniformity of the random numbers obtained from this subroutine depend on an initial input value. To determine input values, sets of 100,000 random numbers were tested for randomness and uniformity. Appendix C explains the four statistical tests used to evaluate the random numbers, the acceptance criteria, and the method developed to randomly

alter the input values to the RANDU subroutine.

Data Generation

When a set of acceptable random numbers (see Appendix C) had been found, this set was used to generate, by use of equation (19), 999 samples of gamma variates for a specific value of skew and sample size. Since the scale parameter, α, was set equal to one, the only variable in equation (19) is the parameter k. By selecting a skew value, the parameter k can be determined since skewness is equal to $1/k^{\frac{1}{2}}$.

Sample sizes of 9, 19, 29, and 49 were selected as these values are of the same order of magnitude as typical values of sample sizes used in flow frequency analysis. For each sample size, skew values of 0.50 (k=4), 0.25 (k=16), and 0.20 (k=25) were used. These skew values were selected on the basis of k being an integer value (Naylor et al., 1966, pp. 87-89) and because they fell within the range of skews considered for this problem, i.e., skew values from zero to one. At a skew value of zero the Pearson Type III distribution is not defined. Rather, a normal distribution exists (see Appendix A). At a skew value of one or larger the Pearson Type III distribution becomes J-shaped (see Appendix A) which is physically meaningless for flow frequency analysis.

For a given sample size and given skew value, a sample was generated using equation (19). To determine if the samples were from a Gamma distribution with the specified

parameters, the sample was tested by the Chi-Square Goodness-of-Fit test and the X^2 (chi-square) value was computed. (Details of this test are given in Appendix C). The sample mean, sample standard deviation, sample skew, and the sample 90 per cent limit were then computed. The sample mean and sample standard deviation were computed by equation (23) and (25) respectively, where the standard deviation is equal to the square root of the variance. The sample skew was computed by equation (26) and the sample 90 per cent limit by equation (27). The ratios of these computed parameters to their known population values were then computed, i.e., $\bar{x}/u$, s/σ, g/G, and S90/P90. For the given value of k, the population values of the mean u, standard deviation, σ, and skew, G, were found from the relationships given with equation (18). The population 90 per cent limit, P90, was found from Karl Pearson's Tables of the Incomplete Gamma Function (1946).

The procedure of generating a sample and computing the ratios of sample to population parameters was repeated 999 times. This gave 999 values of each of the individual parameter ratios. These ratios of sample value to population value were used for ease in analysis since all data were reduced to dimensionless quantities. Also, if the generating technique was good then these ratios should have an expected mean of unity. The 999 values of each ratic (Ratio of Means, Ratio of Standard Deviations, Ratio of Skews, and Ratio of 90 Per Cent Limits) were ranked in ascending order. The mean and

standard deviation of each of the four groups of ratios were computed. The percentage points of each ratio were found by equation (27) and then the data for each ratio were grouped in 21 intervals and the frequency and mid-point of each interval were determined. (For examples of plots of the parameter ratios, see Figures 11, 13, and 14 in Appendix D). This procedure provided the data for the second phase of the research.

Research Phase III

The purpose of the third phase of the research was to determine a set of empirical tolerance factors. This set of tolerance factors was determined directly from the data. From the work of Johnson and Welch (1940), the tolerance factor for any given set of data can be represented by equation (10)

$$k_u = \frac{L - \bar{x}}{s}$$

In equation (10), $\bar{x}$, s, and k_u are random variables and L is a constant related to the proportion, P. In the third phase of the research the distribution of k_u for a Pearson Type III population was determined empirically. For a given skew and given sample size, 999 samples from a transformed Pearson Type III distribution (Gamma distribution) were generated. The generation procedure was the one used in the second phase of

this research. For each sample, a value of k_u was computed by equation (10). (The value of the population limit, L, was computed from tables of percentage points for the Pearson Type III distribution (Harter, 1969) for the given value of the population skewness parameter). This yielded 999 values of k_u which were then ranked in ascending order. From the ranked values of k_u, a value of k_u' could be determined which was larger than epsilon, E, per cent of the other values of k_u. Therefore, E per cent of the time,

$$k_u' \geq k_u$$

and

$$\bar{x} + k_u's \geq \bar{x} + k_us$$

therefore

$$l_u' \geq L \qquad (l_u' = \bar{x} + k_u's)$$

If $l_u' \geq L$ epsilon per cent of the time, then

$$P(l_u' \geq L) = E \tag{28}$$

Equation (28) is an equivalent probability statement to that of the definition of a one-sided upper tolerance limit, equation (6). It is important to note that the tolerance factors developed by the method just described are only applicable for the particular set of data from which they were developed. However, if the 999 samples used to compute the tolerance factors are representative of the population, then the tolerance factors developed from those samples should be representative of the population tolerance factors.

Data Generation

The data (999 samples and hence 999 values of k_u) were generated for sample sizes of 30, 40, 50, 60, 80, and 100. For each sample size, skew values of 0.20, 0.25, 0.50, 0.707, and 1.0 were used. Sets of tolerance factors for a given sample size and skew were computed (equation (10)) for population limit values of L of 90, 95, 99, and 99.9 per cent. For each value of the population limit, L, the tolerance factors corresponding to E values of 0.90, 0.95, 0.99, and 0.999 were determined. Since each set of tolerance factors was based on a set of 999 samples (for a given sample size and skew) and since these factors were intended to serve as guideline estimates of the population values, the data (computed tolerance factors) were "smoothed" by a three part process. This smoothing of data was done because it was reasoned that the population tolerance factors can be represented by smooth, continuous functions. This is the case of the normal

tolerance factors.

Smoothing of Data

The smoothing of the data was done by graphical curve fitting. All curve fitting was done by hand to allow for the weighting of the data, i.e., the tolerance factors for E = 0.90 are more reliable than the factors based on E = 0.999. In a simulation process like the one used for this work, the tails of a distribution are the hardest part of the distribution to define. Another reason for manually fitting curves was to maintain assumed trends in the tolerance factors in regard to continuity, etc..

The purpose of the first part of the smoothing process was to smooth the tolerance factors over the range of probability, E. For a given value of skew and sample size, the tolerance factors for each limit, L, were plotted on probability paper as a function of the E values. Smooth curves were then fit to the data points, see Figure 5.

The purpose of the second part of the smoothing process was to smooth the tolerance factors over the range of the skew. For a given value of the population limit and sample size, the tolerance factors for each probability, E, were read from the curves developed in the first part of the smoothing process (see Figure 5). These adjusted tolerance factors were then plotted versus the values of skew (see Figure 6). At the value of skew equal to zero, the distribution is now normal and the normal tolerance factors were

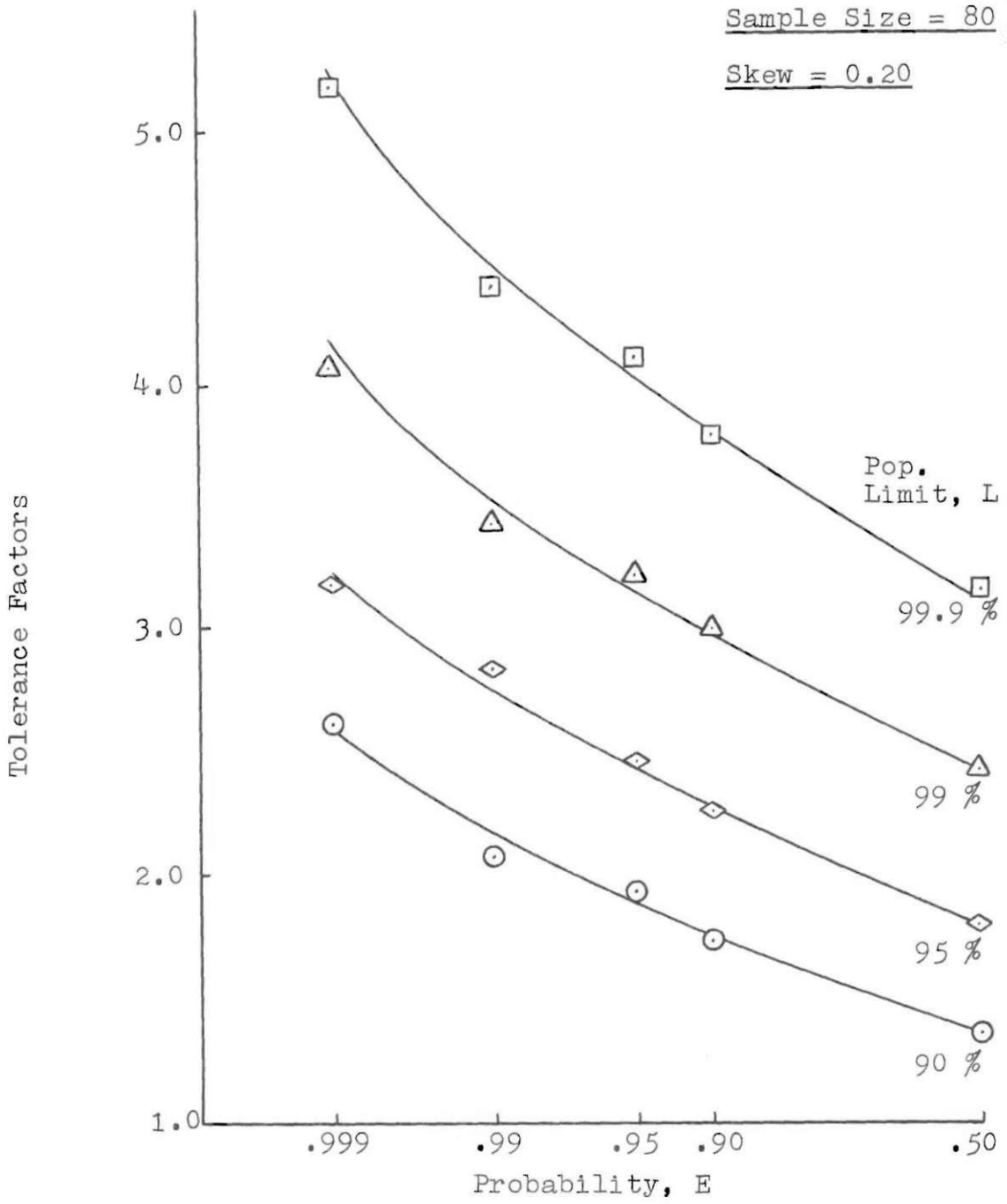


Figure 5. Hypothetical Sample of the First-Part Smoothing Process.

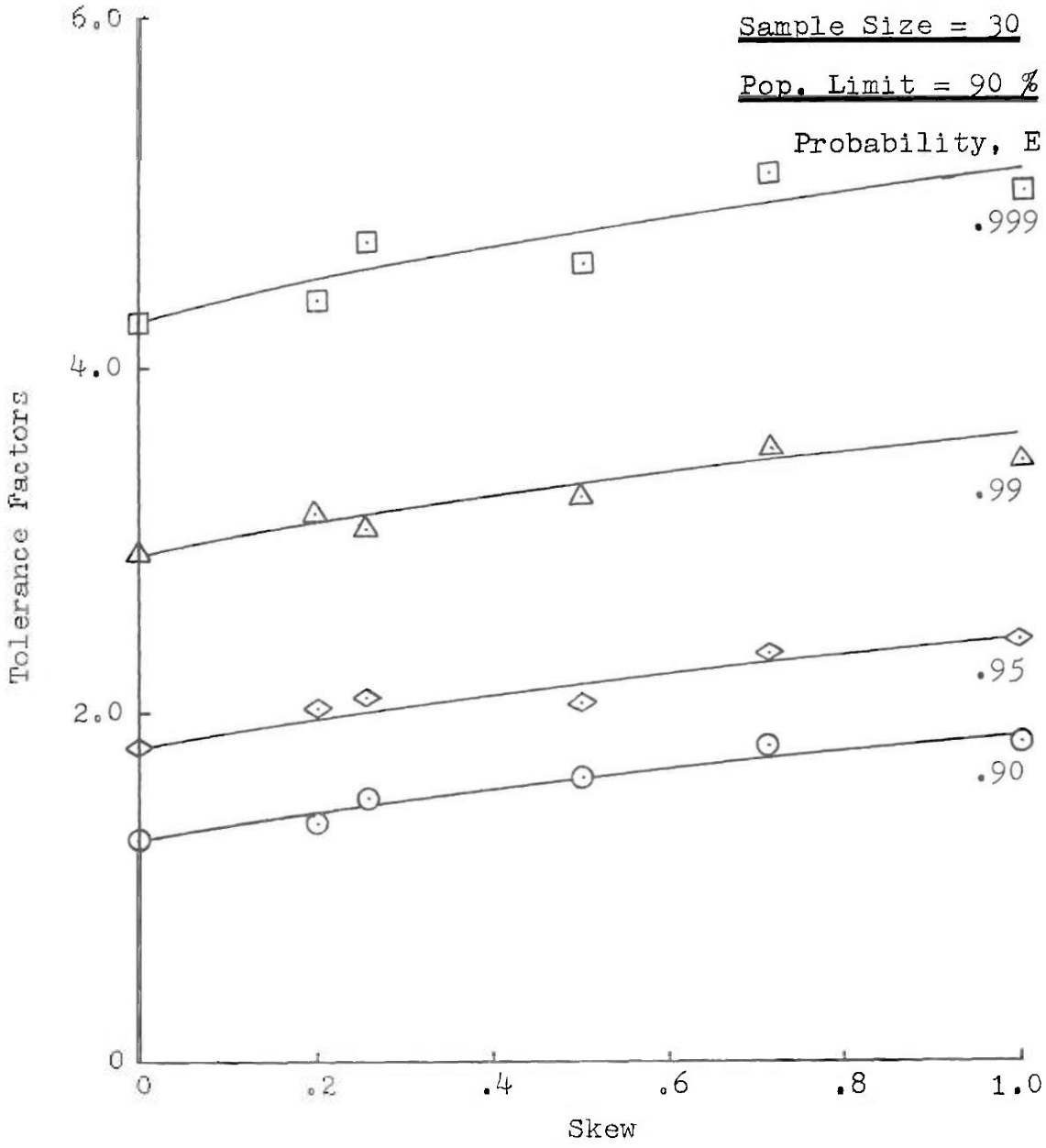


Figure 6. Hypothetical Sample of the Second-Part Smoothing Process.

used. Smooth curves were fitted to the data.

The final step in the smoothing process was designed to smooth the data over the range of the sample size. For a given value of the population limit and skew, values of the tolerance factors for each value of E were read from the curves developed in the second part of the smoothing process (see Figure 6). These tolerance factors were then plotted versus the reciprocal of the sample size. At a sample size of infinity, the population is completely defined, therefore the tolerance factors become the population deviates. At a value of the reciprocal of sample size equal to zero (i.e., sample size equal to infinity) the Pearson deviates obtained from Harter's table (Harter, 1969) were plotted. These values served as a lower bound for the curves that were fitted to the data (see Figure 7).

The values of the tolerance factors were read from the curves developed in the final part of the smoothing process (see Figure 7) and put into tabular form.

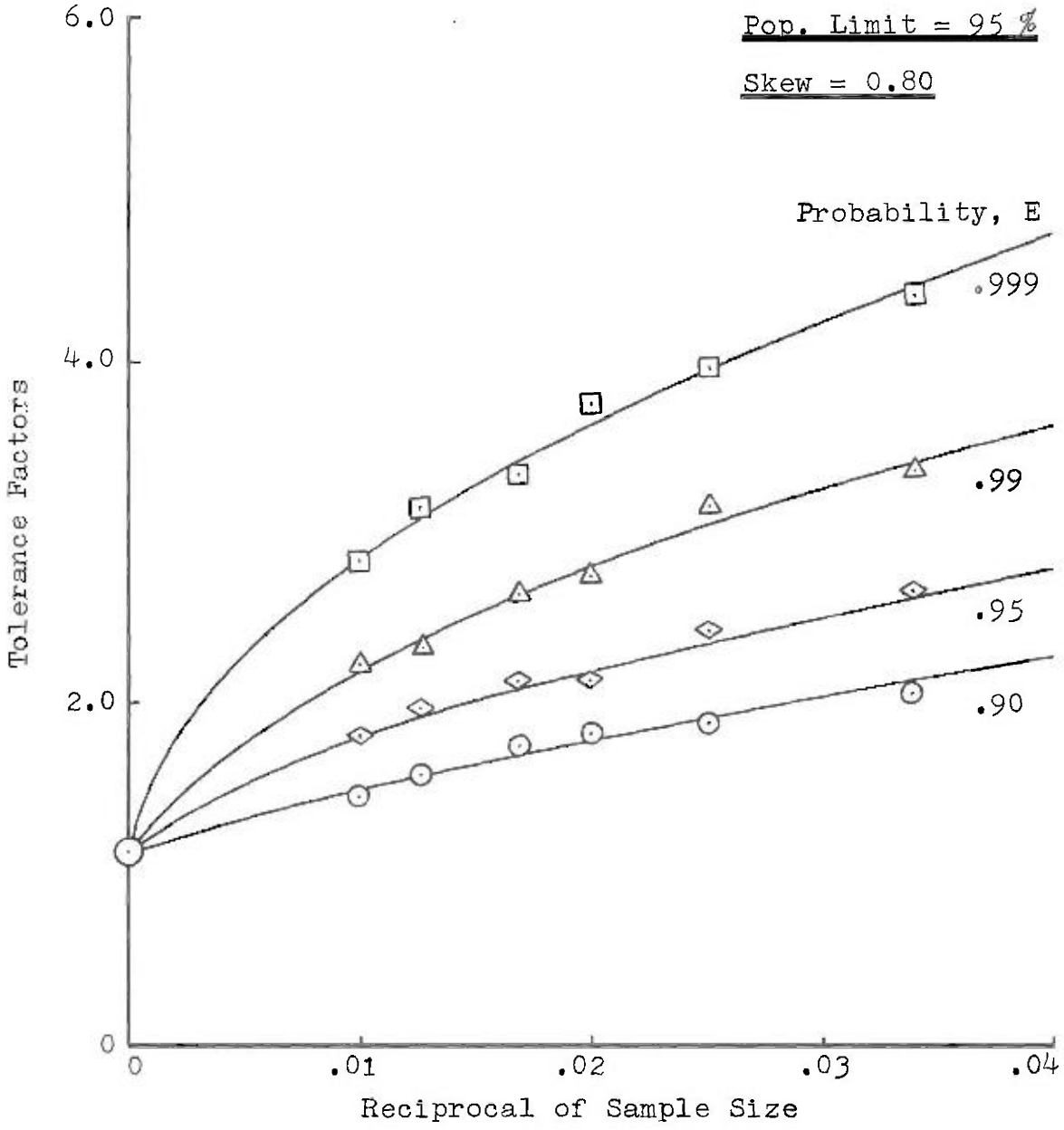


Figure 7. Hypothetical Sample of the Third-Part Smoothing Process.

CHAPTER III

RESULTS

Adequacy of Generated Data

Tests of the data indicate that the generated samples are from a Gamma distribution with the desired parameters. The X^2 (chi-square) values from the Goodness-of-Fit test for each sample were analyzed as being X^2 distributed with (n-1) degrees of freedom, where n is the number of individual X^2 values. A 90 per cent X^2 value was determined from tables of the X^2 distribution (Bowker et al., 1959, pp. 556-557). Theoretically 99.9 sample X^2 values should be greater than the 90 per cent X^2 limit. Table 1 gives the actual number of X^2 values greater than the 90 per cent limit for each sample size and skew value. On the average, the results of the comparison of the number of the theoretical and actual X^2 values are good.

The means and standard deviations of the Ratio of Means, Ratio of Standard Deviations, Ratio of Skews, and Ratio of the 90 Per Cent Limits are given in Table 2. These data indicate that, on the average, the generated samples had the desired values of the mean, standard deviation, and 90 per cent limit. As a further test to determine if the samples were from a Gamma distribution, Karl Pearson's Tables of the Incomplete Gamma Function (1946) was used to compute the 10

Table 1. Results of the Goodness-of-Fit Test of the Generating Technique

Sample Size	Skew	Theoretical Number of X^2 values $\geq$ the 90 per cent limit	Actual Number of X^2 values $\geq$ the 90 per cent limit
19	0.50	99.9	75
19	0.25	99.9	88
19	0.20	99.9	99
29	0.50	99.9	167
29	0.25	99.9	122
29	0.20	99.9	119
49	0.50	99.9	76
49	0.25	99.9	146
49	0.20	99.9	146

Note X^2 values were not computed for sample size = 9.

Table 2. Means and Standard Deviations of the Parameter Ratios (Sample/Population)

Sample Size	Skew	Ratio of Means		Ratio of Std. Dev.	
		Mean	SD	Mean	SD
9	0.20	1.001	.065	.975	.235
9	0.25	1.001	.085	.966	.248
9	0.50	1.007	.121	1.004	.332
19	0.20	.998	.047	.971	.194
19	0.25	.999	.055	.974	.207
19	0.50	1.001	.112	.979	.188
29	0.20	1.000	.040	.991	.141
29	0.25	.998	.047	1.005	.128
29	0.50	.999	.086	1.000	.162
49	0.20	1.000	.030	.998	.093
49	0.25	1.001	.038	.996	.098
49	0.50	1.000	.067	.997	.131

(Continued)

Table 2. Means and Standard Deviations of the Parameter Ratios (Sample/Population) (Continued)

Sample Size	Skew	Ratio of Skews		Ratio of 90 Per Cent Limits	
		Mean	SD	Mean	SD
9	0.20	.359	1.708	1.036	.103
9	0.25	.282	1.424	1.037	.122
9	0.50	.591	.752	1.115	.269
19	0.20	1.222	1.770	1.022	.093
19	0.25	.919	1.536	1.015	.108
19	0.50	.656	.533	1.054	.171
29	0.20	-.020	1.025	1.002	.053
29	0.25	.317	.792	1.011	.062
29	0.50	.703	.536	1.035	.135
49	0.20	-.032	.764	.999	.037
49	0.25	.345	.614	1.006	.051
49	0.50	.853	.415	1.022	.102

through 90 per cent points for a Gamma distribution with a skew of 0.50. Samples (999) with a skew of 0.50 and sample sizes of 29 and 49 were generated and the sample 10 through 90 per cent points were computed by equation (27). The ratios of the sample percentage points to the theoretical percentage points were computed and the means of these ratios were found. The values of the means are shown in Table 3. The data indicate close agreement between sample and theoretical per cent points. The maximum per cent of deviation for a sample size of 29 was 8.1 per cent and for a sample size of 49 was 2.7 per cent.

Generated Samples

Ratio of Means

Sample means theoretically should be approximately normally distributed with mean equal to u and standard deviation equal to $\sigma/n^{\frac{1}{2}}$, where u and σ are the population parameters (Burington et al., 1958, p. 151). For the case of the Ratio of Means, the ratio should be normally distributed with mean equal to unity and standard deviation equal to $(\sigma/n^{\frac{1}{2}})(1/u)$. A comparison of the theoretical mean and standard deviation to the values obtained from the data is shown in Table 4. A plot of the frequency distribution of the Ratio of Means is shown in Figure 11 in Appendix D.

Ratio of Standard Deviations

Theoretically sample means are approximately normally

Table 3. Means of the Ratios of Percentage Points

Skew = 0.50

Percentage Points	Sample Size = 29 Mean of Ratios	Sample Size = 49 Mean of Ratios
10	.91887	.97286
20	.98564	.97602
30	1.00121	.98717
40	1.00366	.99441
50	.99890	1.00040
60	1.02281	1.00642
70	1.02116	1.01329
80	1.02489	1.02007
90	1.03460	1.02240

Table 4. A Comparison of the Numerical and Theoretical Distribution of the Ratio of Means

Sample Size	Skew	Theoretical Mean	Theoretical Std. Deviation	Actual Mean	Actual Std. Deviation
9	0.20	1.0000	0.0666	1.0009	0.0650
9	0.25	1.0000	0.0833	1.0010	0.0852
9	0.50	1.0000	0.1667	1.0073	0.1207
19	0.20	1.0000	0.0459	0.9980	0.0468
19	0.25	1.0000	0.0574	0.9990	0.0546
19	0.50	1.0000	0.1147	1.0009	0.1125
29	0.20	1.0000	0.0371	1.0004	0.0397
29	0.25	1.0000	0.0464	0.9977	0.0473
29	0.50	1.0000	0.0928	0.9989	0.0865
49	0.20	1.0000	0.0286	0.9997	0.0296
49	0.25	1.0000	0.0357	1.0005	0.0377
49	0.50	1.0000	0.0714	1.0004	0.0666

distributed (for large samples), regardless of the underlying distribution from which the samples come. It is not unreasonable therefore, to believe that sample standard deviations from any population may be distributed in approximately one specific manner. The hypothesis that sample standard deviations from a Gamma distribution were distributed in approximately the same manner as those from a normal distribution was tested. For a normal distribution the statistic $(n-1)s^2/\sigma^2$ is distributed as X^2 with $(n-1)$ degrees of freedom. Values of $(n-1)s^2/\sigma^2$ were computed from the data. Plots of the percentage points of $(n-1)s^2/\sigma^2$ and the theoretical X^2 distribution (Figure 12 in Appendix D), and plots of the frequency distribution of the Ratio of Standard Deviations (Figure 13 in Appendix D) indicated that the distribution of $(n-1)s^2/\sigma^2$ might be related to the X^2 distribution. Therefore, Chi-Square Goodness-of-Fit tests were performed for each value of skew for sample sizes of 29 and 49. The theoretical distribution used in the tests for a sample size of 29 was X^2 with 28 degrees of freedom and for a sample size of 49 was X^2 with 48 degrees of freedom. Chi-Square with 48 degrees of freedom is approximately normally distributed with mean equal to 48 and standard deviation equal to $(2(48))^{\frac{1}{2}}$ (Burington et al., 1958, p. 142). The X^2 values for each Goodness-of-Fit test are shown in Table 5. Of the samples tested, only the sets of data for sample size = 29 (skew = 0.25) and sample size = 49 (skew = 0.50)

Table 5. Goodness-of-Fit Test Results for the Distribution of the Sample Variance

Sample Size	Skew	Theoretical 99.9 % X^2 Value	Computed X^2 Value
29	0.20	24.32	17.31
29	0.25	24.32	28.34*
29	0.50	24.32	17.76
49	0.20	24.32	148.27*
49	0.25	24.32	9.19
49	0.50	24.32	22.31

*This set is rejected at the 99.9 per cent significance level.

could be rejected at the 99.9 per cent significance level. Although these results tend to support the hypothesis that $(n-1)s^2/\sigma^2$ for a Gamma distribution is distributed approximately as X^2 with (n-1) degrees of freedom, they are by no means conclusive. To substantiate the proposed hypothesis, a substantial amount of additional data would be required. Also, additional tests would have to be performed on the data besides the X^2 Goodness-of-Fit test used in this work. In view of the amount of time, both man-hour and computer-hour, that would be required for this task, the approach of the numerical evaluation of the distribution of the sample variance for a Pearson Type III distribution was abandoned in favor of the direct determination of empirical one-sided upper tolerance factors, i.e., the third-phase of the research.

Ratio of Skews

The plots of the frequency distributions, see Figure 14 in Appendix D, show the Ratio of Skews to have a unimodal, bell-shaped distribution which is slightly skewed. Of the six frequency distributions of skews, four were skewed right and two were skewed left. However, the distributions are not grouped about the expected value of one. In every case except one (sample size = 19, skew = 0.20) the mean of the distribution was less than unity. A table of the approximate per cent of the distribution which has a value of the Ratio of Skews less than the expected value of unity is Table 6.

Table 6. Percentage of the Distribution of the Ratio of Skews (RSKEW) Below the Expected Mean

Sample Size	Theoretical Skew	% of Distribution with RSKEW less than Unity
9	0.50	65 %
9	0.25	75 %
9	0.20	65 %
19	0.50	75 %
19	0.25	55 %
19	0.20	45 %
29	0.50	75 %
29	0.25	85 %
29	0.20	85 %
49	0.50	65 %
49	0.25	85 %
49	0.20	92 %

Note All percentages are approximate values.

For the sample sizes of 29 and 49, the distribution of the Ratio of Skews shifted away from a value of unity as the population skew value became smaller. Since the tests on the samples to determine if they were from a Gamma distribution with the desired parameters gave good results, it can be assumed that the generated samples would, on the average, have the population skew desired. If this assumption is correct, then the computational form used to compute the sample skew, equation (26), appears to be giving values of the sample skew which are extremely variable and is, in general, underpredicting the population value. The mean of the Ratio of Skews for sample size = 49 (skew = 0.20) underpredicted the expected value of one by 103.2 per cent.

In an effort to explain these results, an attempt was made to determine the effect of the variability of skew in relation to sample size. Matalas and Benson (1968) discussed the standard error of the coefficient of skewness for a normal population as a function of the sample size. Their discussion is based on the work of R. A. Fisher. Fisher (1931) developed an expression for the standard error of the coefficient of skewness* for samples from a normal population.

Using the generating technique previously described, 1000 gamma variates from a population with skew of 0.50 were generated. Selecting various sample sizes from 9 to 1000,

---

*The coefficient of skewness is twice the value of the skew.

the sample mean, standard deviation, and skew were computed. The ratios of these sample values to their population values were also computed. This procedure was repeated using a different set of gamma variates. Table 7 gives the results of these computations. As can be expected, the sample mean and standard deviation converged toward their population values at a greater rate than did the sample skew. A trend toward underprediction of the population skew is apparent. At a sample size of 1000 the sample skew underpredicts the population skew by a minimum of about 11 per cent.

For contrast, 1000 variates were generated from a normal population with mean equal to zero and standard deviation equal to unity. (The generation technique used to generate samples from a normal population is described in Appendix C.) Using the same sample sizes as in the tests on the gamma variates, the sample mean, standard deviation, and skew were computed. The ratios of sample to population value could not be obtained because the population value of both mean and skew is equal to zero. (The results are given in Table 8.) These results indicate that the skew value converges toward the population value at a greater rate than the skew value from a Gamma distribution. Also, the underprediction tendency present for the gamma samples is not apparent for the normal samples. A plot of the sample skew values (expressed in units of deviation from the population value) for the gamma and normal samples as a function of the sample

Table 7. Results of the Test of the Variability of the Sample Skew (Gamma Distribution)

Population Skew = 0.50

Data Set I

Sample Size	Ratio of Means	Ratio of Std. Deviations	Ratio of Skews
9	1.01586	.58179	.45375
19	1.08357	.77933	.60156
29	1.04134	.82442	.36478
49	1.07129	.82585	.21808
75	1.02049	.86241	.18486
100	.99420	.85305	.21734
150	.99601	.87415	.31094
200	1.00072	.85796	.38173
500	.98300	.88884	.68890
750	.97906	.92449	.87649
1000	.97133	.92446	.83633

(Continued)

Table 7. Results of the Test of the Variability of the Sample Skew (Gamma Distribution) (Continued)

Population Skew = 0.50

Data Set II

Sample Size	Ratio of Means	Ratio of Std. Deviations	Ratio of Skews
9	1.09563	.75704	1.07174
19	.93482	.84785	1.05186
29	.92285	.78822	.79521
49	.93558	.88090	.67286
75	.94641	.88746	.70937
100	.96626	.89489	.54808
150	1.00825	.91817	.64019
200	1.01449	.92696	.63254
500	.99171	.93631	.81463
750	.98782	.95002	.86184
1000	.99752	.97621	.88624

Table 8. Results of the Test of the Variability of Sample Skew (Normal Distribution)

Sample Size	Sample Mean	Sample Std. Deviations	Sample Skew
9	-.16402	.78654	.14192
19	-.11811	.87986	.03947
29	-.03163	.92323	.00746
49	-.04372	.94858	-.16439
75	.03721	.90227	-.16685
100	.06518	.90863	-.08834
150	.00549	.90246	-.05189
200	.03158	.91262	-.05687
500	-.02794	.95928	.03807
750	-.03595	.96580	.05112
1000	-.02201	.96030	.03646

Population Mean = 0.0

Population Std. Deviation = 1.0

Population Skew = 0.0

size is shown in Figure 8.

Most of the variability in the skew estimate can be explained by the work of E. S. Pearson. E. S. Pearson (1963) demonstrated the effect of distribution shape on computed sample parameters. To show the effect of different regions of a distribution on the computed moments of that distribution, E. S. Pearson (1963, p. 98) plotted the function

$$c(x) = (x - u)^s f(x)/u_s \qquad (29)$$

where s is the order of the computed moment

u_s is the value of the s^{th} ordered moment about the mean

and f(x) is the density function of the distribution. The c(x) function shows the contribution of a particular x variate to the value of the computed s^{th} moment of a probability distribution. The s^{th} moment about the mean of a distribution is denoted by (Bowker et al., 1959, p. 34)

$$u_s = E(x-u)^s = \int_{-\infty}^{\infty} (x-u)^s f(x)\, dx$$

where $E(x-u)^s$ is called the expected value of $(x-u)^s$. If the function $(x-u)^s f(x)$ were plotted for values of x, the area under the resulting curve would be equal to the

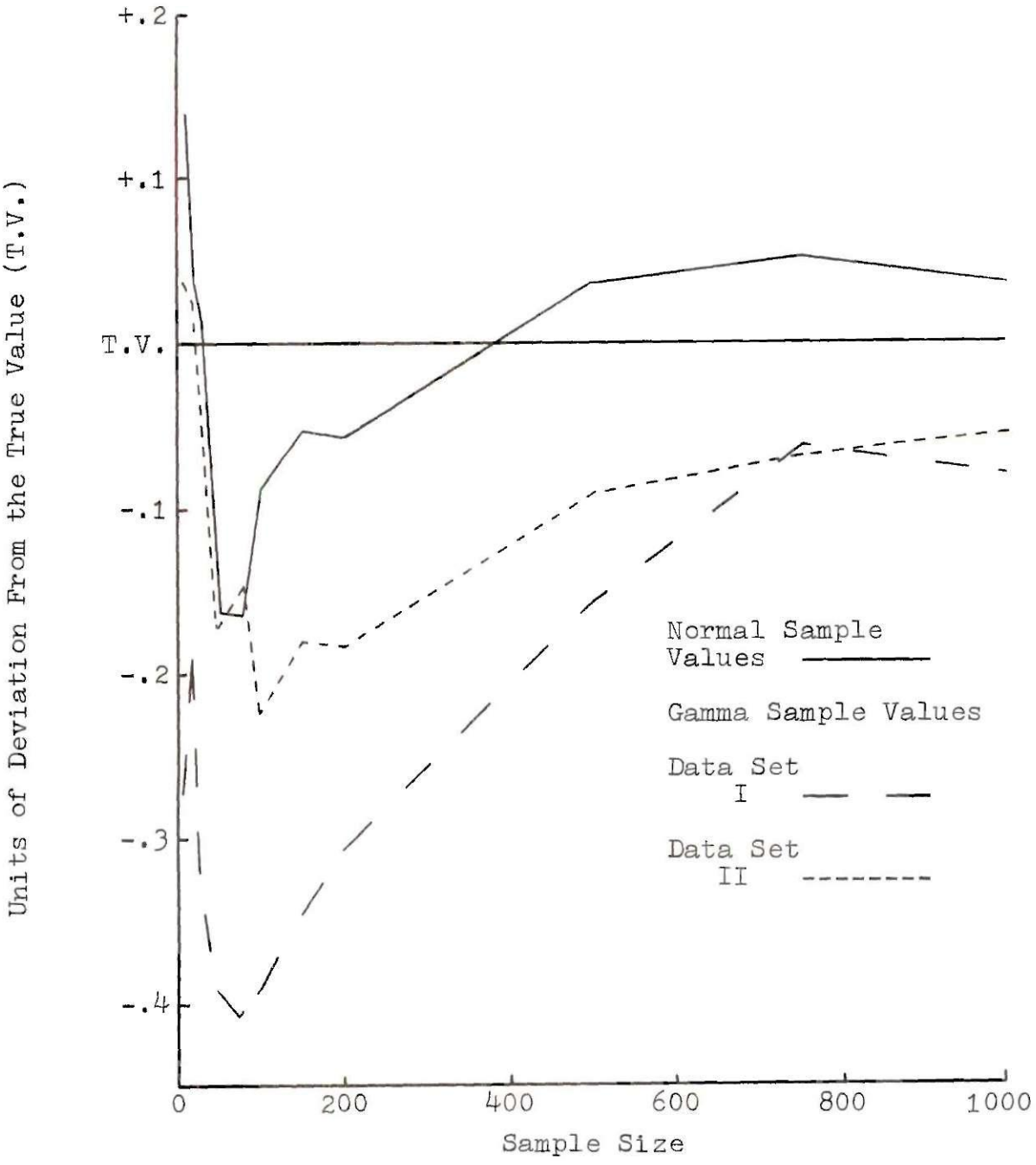


Figure 8. Comparison of the Variability of the Sample Skew for a Normal and Pearson Type III (Gamma) Distribution.

value of the s^{th} moment. This plot can be compared to a plot of the probability distribution, f(x). The area under a plot of the probability distribution is equal to unity. Therefore, for comparison purposes, the area under the plot of the moment function, $(x-u)^s f(x)$, should also be unity. This can be accomplished by dividing the ordinates of the plot of the moment function by the value of the s^{th} moment; which is equivalent to the expression for c(x), equation (29). By comparing the plots of the probability distribution and the c(x) function, the effect of x variates from different regions of the distribution on the value of the s^{th} moment can be observed.

Equation (29) has been evaluated for the third-order moment, which is the one used in the computation of the sample skew (see equation (26)). The density function, f(x), represents the Gamma distribution with skew equal to 0.50. The c(x) function, as well as the density function, f(x), is shown in Figure 9. Figure 9 illustrates that the value of the third moment depends largely on variates from the tail of the distribution. The extreme tail of a distribution is the hardest portion of the distribution to accurately define by simulation (i.e., with samples). This fact would account for the variability in the skew estimate. Also, because the long tail contributes to the "positiveness" of the computed third moment, a failure to define the long tail will result in a loss of "positiveness" in the estimated value. Hence, the

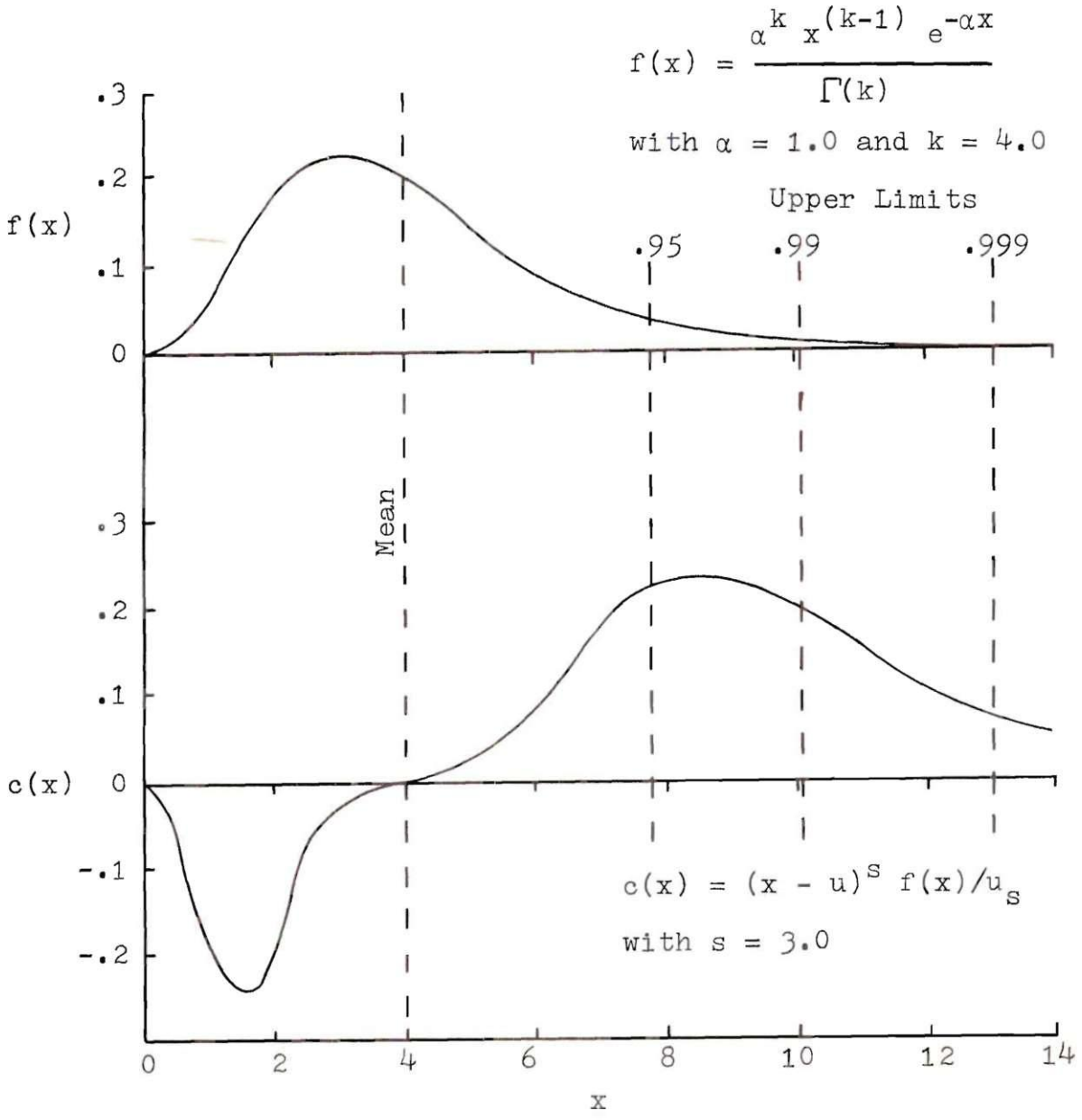


Figure 9. Distribution of f(x) for a Gamma Distribution and c(x) for the Third Moment.

result would be underpredicting the true value, a result verified by this research data.

For the case of the normal distribution, the contribution to the third moment is the same in both tails because of the symmetry of the normal distribution. Therefore, the tendency to underpredict the population value of the third moment should not occur, another result verified by this research data.

The findings of this research, although based on a limited amount of data, indicate that a computed skew value can have great variability as well as inherent prediction tendencies. These findings are important in regard to flow frequency analysis. One method used in flood frequency analysis consists of computing a skew value from a sample of flood data. The computed skew value is then used to determine the distribution (i.e., the Pearson Type III curve with appropriate parameters) to represent the data (Bulletin No. 15, 1967). However, the extreme tail of a distribution is the hardest portion to define with a sample. Therefore, the computed skew value used to determine the distribution of flood frequencies, may exhibit the characteristics of variability and underprediction as previously noted.

The use of a regionalized skew, averaged over a number of skew values (Bulletin No. 15, 1967, p. 13) may be more reliable than single sample estimates. However, the use of an averaged skew does not necessarily eliminate the

chance of underprediction. Also, as indicated by Figure 8, caution should be exercised in extrapolating the normal skew coefficient standard errors of Matalas and Benson (1968) to the Pearson Type III distribution.

Empirical Tolerance Factors

One-sided upper tolerance factors developed from the simulated data are presented in Table 10 in Appendix D. These factors are given for population skew values of 0.20, 0.40, 0.60, 0.80, and 1.0. These tolerance factors are larger than normal tolerance factors for the same sample size. Also, the empirical tolerance factors exhibit the required characteristics of tolerance factors. The factors increase as skew, probability, or proportion (limit) increase, and decrease as sample size increases.

These tolerance factors were developed to demonstrate a method of empirically determining tolerance factors from given data. These factors are only estimates of the population one-sided upper tolerance factors and are intended only as guideline values for future work. Use of these factors require that a sample come from a distribution which is known to be distributed as a Pearson Type III curve with a known skew value. Also, the sample standard deviation must be computed by equation (25).

CHAPTER IV

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

This work has illustrated the role that simulation on a digital computer can play in hydrologic studies. Based on the test criteria used for this research, the Chi-Square (X^2) Goodness-of-Fit test and the average value of the ratio of sample to population parameters, the generating technique used is an acceptable method of generating samples from a Pearson Type III distribution.

The distribution of $(n-1)s^2/\sigma^2$ for a Pearson Type III population could not be conclusively shown to be distributed as X^2 with $(n-1)$ degrees of freedom. This conclusion is based on the result of six X^2 Goodness-of-Fit tests with a 99.9 per cent significance level. Also, the determination of the distribution of a function of the sample variance for a Pearson Type III population by direct mathematical means will be very difficult.

The method of direct estimation of one-sided upper tolerance factors from data yielded a set of factors which estimate the values of the actual tolerance factors for a Pearson Type III distribution. This conclusion is based on the fact that the data used to develop these tolerance factors

has been statistically accepted as being from a Pearson Type III distribution. These empirical tolerance factors are intended to serve as guideline values for future work.

Recommendations

The empirical tolerance factors developed in this research can serve as a basis for future work in two major directions. First, these factors may aid in the analytical development of one-sided upper tolerance factors for a Pearson Type III or Gamma distribution. Although the results of this work could not prove that $(n-1)s^2/\sigma^2$ from a Pearson Type III population was distributed as X^2 with $(n-1)$ degrees of freedom, they certainly indicated that this is approximately the case. Additional work on this hypothesis may result in its verification. If the distribution of $(n-1)s^2/\sigma^2$ from a Pearson Type III population can be shown to be X^2 distributed with $(n-1)$ degrees of freedom, then a similar statistic to the non-central "t" statistic will probably represent the distribution of the population Pearson Type III tolerance factors (see pages 15-19). In any case, regardless of the type of statistic developed, these empirical tolerance factors can serve as guidelines to the form and characteristics the actual tolerance factors must possess. They can also serve as a comparison to any theoretically developed factors.

The second way these empirical tolerance factors may be used is as a means of obtaining practical or design

factors, without having to determine theoretical values. To accomplish this, the method of determining empirical factors developed for this research could be used. Although the three-part smoothing technique performed on this data was done manually, this procedure could probably be represented in mathematical terms. A numerical weighting technique would have to be developed to reflect the reliability of different parts of the empirical distributions. If this mathematical representation can be made, the entire procedure, including generation and smoothing of data, can be done on the digital computer. Then a number of sets of tolerance factors could be developed and the factors could be averaged over all sets. The resulting values should be estimates of the actual values that are reliable enough for practical application.

Finally, if a computed sample skew value is to be used as a determining parameter, for example, in curve fitting or the selection of tolerance factors, then additional work needs to be done in defining the variability of the skew estimate. The standard error of the computed skew or skew coefficient as a function of sample size must be known for a Pearson Type III distribution. The standard error of the coefficient of skewness for a Pearson Type III distribution has not been developed for sample sizes less than approximately 100 (Matalas et al., 1968). If the standard error can not be developed analytically, then simulation techniques might be used to develop the standard error of the coefficient of

skewness for the Pearson Type III distribution.

A P P E N D I C E S

APPENDIX A

THE PEARSON TYPE III DISTRIBUTION

Development

Karl Pearson derived a series of probability functions to fit virtually all frequency distributions. These functions were developed by first selecting a mathematical expression to represent all types of frequency curves. This mathematical expression had to satisfy two conditions. First, the curve must be tangent to the x-axis at at least one end, that is, when $y = 0$, $dy/dx = 0$. Second, the curve must have a maximum, that is, at some value of x such as $x = -a$, then $dy/dx = 0$. These two conditions are satisfied by the following expression

$$\frac{dy}{dx} = \frac{y(x + a)}{F(x)} \tag{30}$$

where $F(x)$ is any function of x.

By Maclaurin's theorem, $F(x)$ can be expanded such that $F(x)$ can be represented by, $F(x) = b_0 + b_1x + b_2x^2 + \ldots$. For practical purposes, there is no need to consider the terms past b_2x^2. Equation (30) can now be written

$$\frac{dy}{dx} = \frac{y(x + a)}{b_0 + b_1x + b_2x^2}$$

Rearranging terms yields

$$\frac{dy}{y} = \frac{(x + a)}{b_0 + b_1x + b_2x^2}\, dx \qquad (31)$$

Integration of equation (31) gives

$$\ln y = \int \frac{(x + a)}{b_0 + b_1x + b_2x^2}\, dx \qquad (32)$$

Taking the antilogarithm (to the base e) of both sides of equation (32) gives the expression

$$y = e^{\int \frac{(x + a)}{(b_0 + b_1x + b_2x^2)} dx} \qquad (33)$$

Equation (33) is the general Pearsonian equation for all frequency curves, where a, b_0, b_1, and b_2 are constants (Elderton et al., 1969, p. 41). These constants (a, b_0, b_1, and b_2) can be expressed in terms of the first four moments

about the mean. (The first moment about the mean, u_1, is equal to zero.) This yields three new terms, β_1, β_2, and K, which are defined as follows (Elderton et al., 1969, p. 45)

$$\beta_1 = u_3^2/u_2 \tag{34}$$

$$\beta_2 = u_4/u_2^2 \tag{35}$$

$$K = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6)} \tag{36}$$

where u_2, u_3, and u_4 are the second, third, and fourth moments about the mean.

The values of the terms β_1, β_2, and K are used as criteria to determine the different types of Pearsonian distributions. The development of β_1, β_2, and K as well as the development of the different types of Pearson distributions has been documented by W. P. Elderton (1969, pp. 35-109).

Form and Parameters

For the curve of interest in this research, the Pearson Type III, the value of K is equal to infinity and β_1 and β_2 must satisfy the relationship, $2\beta_2 - 3\beta_1 - 6 = 0$. The Pearson Type III curve has the form (Elderton et al., 1969, pp. 78-79)

$$y = y_o(1 + x/a)^p \, e^{-\gamma x} \qquad -a \leq x < \infty \qquad (37)$$

with $p = \gamma a$

$$y_o = \frac{p^{p+1}}{ae^p \Gamma(p+1)}$$

where p is the skewness parameter

a is the lower bound of the curve

y_o is the value of the curve at the mode, i.e., at $x = 0.0$

and $\Gamma(p+1)$ is the complete gamma function.

The Pearson Type III curve is limited in one direction and skewed. The origin of the axis is at the mode. The curve parameters, p, γ, and a can be expressed in terms of the moments about the mean (Elderton et al., 1969, p. 78)

$$\gamma = 2u_2/u_3$$

$$p = (4u_2/u_3^2) - 1 = (4/\beta_1) - 1$$

$$a = (2u_2^2/u_3) - (u_3/2u_2)$$

The statistical parameters of the curve are determined by (Pearson, 1946, pp. vii - viii)

$$\text{Mode} = \text{Mean} - (u_3/2u_2)$$

$$\text{Mean} = (p + 1)/\gamma$$

$$\text{Variance} = (p + 1)/\gamma^2 = u_2$$

$$\text{Skewness} = \frac{\text{Mean - Mode}}{\text{Std. Dev.}} = 1/(p+1)^{\frac{1}{2}} = \beta_1^{\frac{1}{2}}/2$$

Characteristics

The Pearson Type III curve is bell-shaped (see Figure 2, page 9) for skew values between zero and unity, i.e., for $p > 0$. The curve becomes J-shaped for skew values greater than or equal to unity, i.e., for $p \leq 0$. (Elderton et al., 1969, p. 79.) At a value of the skew equal to zero, the Pearson Type III curve no longer exists. A curve with zero skew can not be limited in only one direction and must have a K value (equation (36)) of zero. Also, the β_1 term (equation (34)) must be equal to zero. These conditions are satisfied by the normal distribution. The normal curve has a skew of zero, is unlimited in both directions, and has the value of the criteria terms, $\beta_1 = 0$, $\beta_2 = 3$, and $K = 0$ (Elderton et al., 1969, p. 71).

When the third moment about the mean, u_3, is negative, then γ and a are negative and the Pearson Type III

curve is now limited at a distance "a" after the mode (Elderton et al., 1969, p. 79). If γ and a are negative the equation for the Pearson Type III curve becomes

$$y = -y_o(1 - x/a)^p \, e^{\gamma x} \qquad -\infty < x \leq a$$

For a negative third moment, the Pearson Type III curve is rotated 180 degrees about its mode and also rotated 180 degrees about the x-axis. Figure 10 represents Pearson Type III curves for positive and negative value of the third moment, u_3.

A negative value of u_3 implies negative skewness (see the computation of skewness, equation (26), page 27). A negative skewness is sometimes encountered in fitting flood data to a Pearson Type III curve for flood frequency analysis.

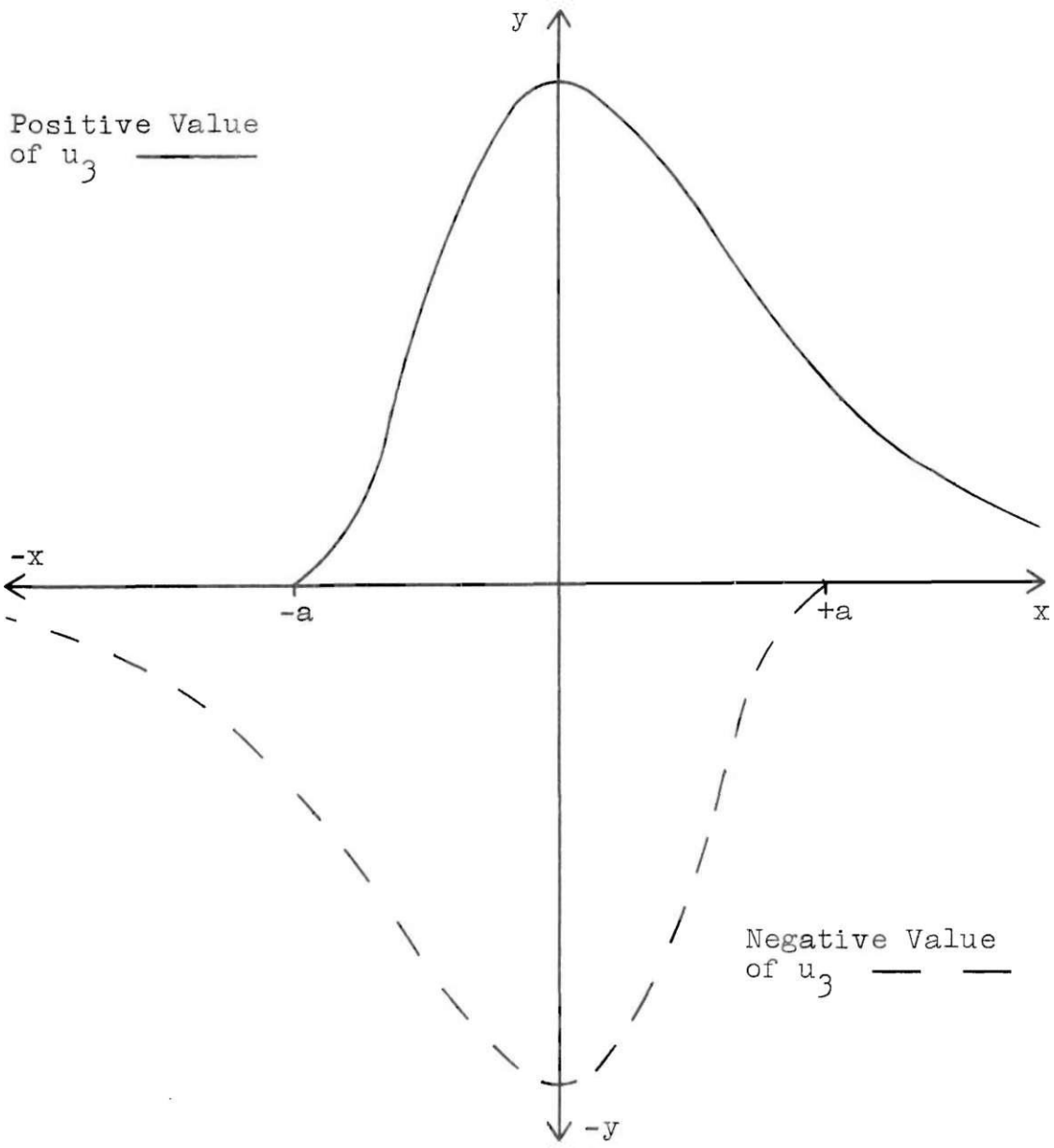


Figure 10. The Pearson Type III Distribution for Positive and Negative Values of the Third Moment.

APPENDIX B

PEARSON TYPE III —— GAMMA TRANSFORMATION

The density function of the Pearson Type III curve is given by equation (37) in Appendix A. To find the area under the Pearson Type III curve, this density function must be integrated over the range of x. This integration yields the cumulative distribution function, F(x)

$$F(x) = \int_{-a}^{x} y_o(1 + x/a)^p \, e^{-\gamma x} \, dx \qquad -a \leq x < \infty \qquad (38)$$

In its present form, the right-hand side of equation (38) can not be integrated. To evaluate this integral, Karl Pearson developed the Tables of the Incomplete Gamma Function (1946). Pearson (1946) transformed the Pearson Type III distribution to the Gamma distribution and then evaluated the Gamma distribution. To transform a Pearson Type III distribution to a Gamma distribution it is necessary only to change the variable of integration.

The expression for a Pearson Type III curve with the origin of the x-axis at the mode is

$$y = y_o(1 + x'/a)^p \, e^{-\gamma x'} \qquad -a \leq x' < \infty \qquad (39)$$

with $p = \gamma a$

$$y_o = \frac{p^{p+1}}{ae^p \Gamma(p+1)}$$

To move the origin of the curve (the lower bound) to zero, Pearson (1946, p. vii) defined a new variable of integration

$$v = p(1 + x'/a) = \gamma(a + x') \tag{40}$$

Letting $x = (a + x')$ and substituting into equation (40)

$$v = \gamma x \qquad 0 \leq v < \infty \tag{41}$$

Substituting the new variable v into equation (39)

$$y = y_o \frac{e^p}{p^p} v^p e^{-v} \tag{42}$$

Substituting the expression for y_o into equation (42) and simplifying terms yields

$$y = \frac{\gamma}{\Gamma(p+1)} v^p e^{-v} \tag{43}$$

The complete gamma function is defined by (Pearson,

1946, p. v)

$$\Gamma(p+1) = \int_0^{\infty} e^{-x} x^p \, dx \qquad (44)$$

This function has been evaluated and put in tabular form for various values of the argument p (Selby, 1965, p. 349). The incomplete gamma function is defined by (Pearson, 1946, p. v)

$$\Gamma_x(p+1) = \int_0^{x} e^{-x} x^p \, dx \qquad (45)$$

If equation (43) is integrated from zero to v, then the right hand side of the resulting expression will be a combination of the complete and incomplete gamma functions. By evaluating the incomplete gamma function (equation (45)), Karl Pearson (1946) was able to evaluate the cumulative distribution function of the Gamma distribution.

For the purpose of this research, it was desirable to have the Gamma distribution in terms of the variable x ($x = (a + x')$). Therefore, substituting the value of v, i.e., γx (equation (41)) into equation (43) yields

$$y = \frac{\gamma^{p+1}}{\Gamma(p+1)} x^p e^{-\gamma x} \qquad 0 \leq x < \infty \qquad (46)$$

From Computer Simulation Techniques (1966) the form of a Gamma distribution is given by (Naylor et al., 1966, pp. 87-89)

$$y = \frac{\alpha^k \, x^{(k-1)} \, e^{-\alpha x}}{\Gamma(k)} \qquad 0 \leq x < \infty \tag{47}$$

where α is the scale parameter

k is the shape or skewness parameter

and $\Gamma(k)$ is the complete gamma function which is equal to $(k-1)$ factorial for integer values of k.

If the parameters in equation (46) are redefined such that $\alpha = \gamma$ and $k = (p+1)$, then equation (46) is identical to equation (47).

Gamma Parameters

The statistical parameters of the Gamma distribution can be determined by the expressions

$$\text{Mean} = k/\alpha \tag{48}$$

$$\text{Variance} = k/\alpha^2 \tag{49}$$

Pearson (1946, p. viii) shows the expression for the skewness to be

$$\text{Skewness} = 1/(p+1)^{\frac{1}{2}}$$

Therefore, in the notation of equation (47)

$$\text{Skewness} = 1/(k)^{\frac{1}{2}}$$

The second and third moments about the mean expressed in terms of α and k are (Kendall et al., 1963, p. 62)

$$u_2 = k/\alpha^2$$

$$u_3 = 2k/\alpha^3$$

Relationship of Gamma Parameters to the Tolerance Factor

Since only k determines the shape of the Gamma distribution, Pearson (1946, p. vii) chose to express the integral of the Gamma distribution in terms of the shape parameter alone. The value of the limit (corresponding to a specified probability level) of a Gamma distribution with a particular skew value can be read from Pearson's tables. To determine the value of that limit for a Gamma distribution (with the same skew parameter) with a value of the scale parameter, α, other than unity, the limit can be computed by

$$L_{k,\alpha} = L_k/\alpha \tag{50}$$

where $L_{k,\alpha}$ is the limit for a Gamma distribution with parameters k and α

L_k is the limit for a Gamma distribution with parameter k, read from Pearson's tables

and α is the scale parameter.

One-sided upper tolerance factors can be defined by equation (10) (page 16)

$$k_u = \frac{L - \bar{x}}{s}$$

For any Gamma distribution with parameters α and k, the limit L is defined by equation (50) and the values of $\bar{x}$ and s can be written in terms of α and k (equations (48) and (49)). Substituting these values into equation (10) gives

$$k_u = \frac{(L_k/\alpha) - (k/\alpha)}{k^{\frac{1}{2}}/\alpha} = \frac{L_k - k}{k^{\frac{1}{2}}} \tag{51}$$

From equation (51), the one-sided upper tolerance factor is independent of the scale parameter, α.

APPENDIX C

COMPUTER TECHNIQUES

Random Number Generation

Random numbers uniform on the interval from zero to one, U(0,1), were required for the generation of samples from a Gamma distribution (equation (19), page 23). Although tables of random numbers are available, they were not used. The research required in excess of one million random numbers so the random numbers were not input to the computer. Rather, the required random numbers were generated on the computer. The computer system used, the UNIVAC 1108, had subroutines available which use mathematical relationships to produce pseudorandom numbers. Pseudorandom numbers are numbers for which a hypothesis of randomness can not be rejected within specified statistical limits. The technique used to generate random numbers had two requirements. First, each set of random numbers should possess the qualities, within certain acceptable levels, of randomness and uniformity. Second, different sets of random numbers should be independent of each other.

Random numbers U(0,1) were obtained from the RANDU subroutine in the Math-Pack (a group of subroutines on the UNIVAC 1108 system). Generation of random numbers by the

RANDU subroutine requires an initial input integer value. Again because of the quantity of random numbers required, a single set of random numbers was not generated. Instead, various sets of random numbers were generated as they were required in the program to produce the gamma variates. A technique was developed to vary the initial input value to the RANDU subroutine in order that each of the sets of random numbers obtained was independent of all others.

Three techniques to vary the input value to the RANDU subroutine were tried. For each technique, 100,000 random numbers U(0,1) were generated and tested for uniformity and randomness. Four statistical tests were used. A 95 per cent significance level was used in each test. The results of these tests determined which of the three techniques of input number variation would be used in the generation of samples from a Gamma distribution.

The random numbers were obtained by generating 100 sets of 1000 numbers each. Each of the 100 sets was tested by the Lagged Product test, the Test of Runs, and the Frequency test. The largest random number in each of the 100 sets was obtained and these numbers were tested by the Maximum test. Also, the X^2 (chi-square) values obtained from the Frequency test on each of the 100 sets were tested using a Frequency test of X^2 values. All of these tests are described in Computer Simulation Techniques by Naylor et al., (1966, pp. 57-62).

The first technique of input value variation consisted

of selecting odd* integer values from a random number table (Selby, 1965, pp. 251-257). These random integers were then used as the input values to the RANDU subroutine.

For the second technique, an initial odd integer was used as input to generate one set of random numbers U(0,1). The last number in this generated set was made larger than unity by multiplying it by factors of ten. This value was then rounded off to make it an integer, multiplied by two to make it even, and added to the initial chosen odd integer. The result was used as input to the RANDU subroutine, a set of random numbers U(0,1) was generated, and then the above procedure of computing an input value from the last random number in the set was repeated.

The final technique of input value variation involved the use of a computer subroutine that produced pseudorandom integers, uniform on the interval from zero to 2^{27}. The NRAND subroutine in the Math-Pack was used to obtain the pseudorandom integers. Each integer was divided by 100 because of a integer size limitation in the RANDU subroutine. This value was multiplied by two to make it an even integer and added to five to make it odd. These adjusted random integers were then used as input values to the RANDU

---

*Odd input values were used for the random number generation because of the nature of the mathematical relationships of pseudorandom number generation (Naylor et al., 1966, pp. 47-57).

subroutine. To eliminate any initial cycling* in the random number generaters (NRAND and RANDU), the first 250 random integers and the first 200 random numbers U(0,1) produced by these subroutines were not used.

The results of the statistical tests of the random numbers U(0,1) indicated that the procedure of generating random integers (NRAND subroutine) and in turn using these integers (adjusted to make them odd) as input values to the RANDU subroutine was the best of the three techniques tried for input value variation. This technique was selected for use in the generation of samples from a Gamma distribution.

The NRAND subroutine that produced the random integers requires two initial input values. The ones used for this work were I = 3 and J = 97473779. The results of the Lagged Product test, the Maximum test, and the Frequency test of the X^2 values, for the selected technique of random number generation are given in Table 9.

Test of Goodness of Fit

The Chi-Square Goodness-of-Fit distribution has been discussed by Markovic (1965, pp. 10-15). For the Goodness-of-Fit test either intervals of equal length or intervals of equal probability can be used. For this research, intervals of equal probability were selected.

---

*Sometimes the initial numbers obtained from a pseudo-random generator are considerably less random than those numbers obtained after the initial numbers have been generated.

Table 9. Results of the Tests of Random Numbers

Test	Actual Value		Theoretical Value or Rejection Level	
*Lagged Product Test	Mean	Std Dev	Mean	Std Dev
	-.076	1.094	0	1.0
Maximum Test	Computed X^2 (9 d.f.**)		95 % X^2_9 value	
	3.400		16.919	
Frequency Test of X^2 Values	Computed X^2 (9 d.f.)		95 % X^2_9 value	
	12.000		16.919	

*The mean and standard deviation shown in this table represent the mean and standard deviation of the normal variates computed in the 100 individual Lagged Product tests.

**d.f. denotes degrees of freedom.

The Stat-Pack* contained a subroutine (GAMIN) that calculated the value of the incomplete Gamma distribution. At a value of X, this subroutine gives the corresponding cumulative probability of the incomplete Gamma. No inverse exists for the incomplete Gamma distribution, therefore, an iterative procedure was used to determine the equal probability intervals. The restrictions on the Chi-Square Goodness-of-Fit test require that the theoretical frequency in each interval be at least five (Markovic, 1965, p. 10). If the frequency, f, in each interval is set, the probability, p_i, required in that interval can be computed, i.e., $p_i = n/f$, where n is the sample size.

For this work, the theoretical frequency in each interval was set equal to six and the required probability computed for a given sample size. An initial small value of X was selected and the cumulative probability obtained from the GAMIN subroutine. The value of X was then incremented until the value of the cumulative probability from the GAMIN subroutine equaled the required interval probability, p_i. The value of X at which the cumulative probability equaled the interval probability then became the interval limit. The X value of the interval limit was then incremented until the value of the cumulative probability from the GAMIN subroutine equaled the required probability in the first two intervals.

---

*A group of subroutines on the UNIVAC 1108 system.

This corresponding X value became the second interval limit. This procedure was repeated until a value of the cumulative probability of unity was reached.

The iterative procedure of determining intervals of equal probability is not an exact method. Therefore, a balancing routine was used if the difference in the frequency (np_i) in the last two intervals was greater than one. The balancing routine consisted of determining the excess probability (the difference in frequency in the last two intervals divided by the sample size) and distributing it throughout all the intervals. The interval limits (the X limit values) were increased so as to include the required amount of excess probability, p_e, in each interval. Since the intervals were determined for equal probability, they are not necessarily of equal length. Therefore, the interval limits were increased in proportion to ratio of the increase in cumulative probability to the increase in the X value. Mathematically the increase in each interval limit was determined by

$$X'(i) = X(i) + \left(\frac{X(i+1) - X(i)}{CP(i+1) - CP(i)}\right) \frac{p_e}{NI}$$

where $X'(i)$ is the limit after balancing

$X(i)$ is the orginally computed interval limit for the i^{th} interval

$X(i+1)$ is the limit for the $i^{th} + 1$ interval

CP(i) is the cumulative probability at the i^{th} interval limit

CP(i+1) is the cumulative probability at the $i^{th} + 1$ interval limit

p_e is the excess probability to be distributed

and NI is the number of intervals.

Once the intervals had been determined, the theoretical frequency existing in each interval was computed by multiplying the probability in each interval by the sample size. With the intervals and theoretical frequency in each interval determined, the actual frequency in each interval was obtained from the generated data and the Chi-Square Goodness-of-Fit test performed.

Normal Sample Generation

Samples from a normal distribution were generated through the use of the inverse of the cumulative normal distribution. The cumulative normal distribution has a probability range from zero to one. The Stat-Pack contains a subroutine (TINORM) which gives the value of the inverse of the normal distribution for a given probability value. The value of the inverse of the normal distribution is given in terms of a standardized deviate, z; where $z = (x-u)/\sigma$, and x is a normal variate and u and σ are the population mean and standard deviation.

To generate samples from a normal distribution, a set

of random numbers U(0,1) was first obtained in the same manner as those obtained for Gamma sample generation. These random numbers U(0,1) were used to represent probability values and were input to the TINORM subroutine. From the random probability values, the TINORM subroutine gave the corresponding random normal standardized deviates. These standardized deviates, z, were used to produce the normal variates, x, from a distribution with a specific mean, u, and standard deviation, σ. The normal variates were computed by

$$x = \sigma z + u$$

APPENDIX D

TABLES AND ILLUSTRATIONS

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution

Skew = 0.20

Sample Size	Probability, E = 0.90				Probability, E = 0.95			
	Proportion				Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	1.79	2.27	3.34	4.68	1.93	2.45	3.57	4.98
40	1.72	2.20	3.24	4.54	1.84	2.34	3.42	4.79
50	1.67	2.15	3.17	4.45	1.78	2.27	3.33	4.66
60	1.63	2.11	3.13	4.39	1.74	2.23	3.26	4.58
70	1.61	2.09	3.09	4.33	1.70	2.19	3.21	4.51
80	1.59	2.06	3.05	4.28	1.67	2.15	3.17	4.44
90	1.57	2.04	3.02	4.24	1.65	2.13	3.13	4.39
100	1.55	2.02	3.00	4.20	1.63	2.10	3.10	4.34
∞	1.32	1.75	2.62	3.67	1.32	1.75	2.62	3.67

(Continued)

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution (Continued)

Skew = 0.20

Sample Size	Probability, E = 0.99				Probability, E = 0.999			
	Proportion				Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	2.19	2.76	4.01	5.67	2.60	3.25	4.67	6.60
40	2.07	2.61	3.80	5.35	2.40	3.04	4.34	6.08
50	1.99	2.52	3.67	5.15	2.27	2.91	4.14	5.75
60	1.93	2.45	3.58	5.01	2.18	2.81	3.99	5.54
70	1.88	2.40	3.51	4.90	2.11	2.73	3.88	5.38
80	1.84	2.35	3.44	4.80	2.05	2.67	3.79	5.25
90	1.81	2.31	3.39	4.73	1.99	2.61	3.72	5.13
100	1.78	2.27	3.34	4.66	1.95	2.56	3.65	5.05
∞	1.32	1.75	2.62	3.67	1.32	1.75	2.62	3.67

(Continued)

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution (Continued)

Skew = 0.40

Sample Size	Probability, E = 0.90 Proportion				Probability, E = 0.95 Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	1.86	2.45	3.79	5.58	2.01	2.64	4.08	6.00
40	1.78	2.36	3.66	5.42	1.91	2.52	3.90	5.74
50	1.73	2.31	3.58	5.30	1.85	2.44	3.78	5.58
60	1.69	2.27	3.52	5.22	1.80	2.39	3.70	5.44
70	1.66	2.24	3.47	5.15	1.77	2.35	3.63	5.36
80	1.64	2.21	3.43	5.09	1.74	2.31	3.57	5.28
90	1.62	2.19	3.39	5.04	1.71	2.28	3.53	5.22
100	1.60	2.17	3.36	5.00	1.69	2.25	3.48	5.15
∞	1.34	1.84	2.89	4.24	1.34	1.84	2.89	4.24

(Continued)

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution (Continued)

Skew = 0.40

Sample Size	Probability, E = 0.99				Probability, E = 0.999			
	Proportion				Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	2.29	2.99	4.57	6.80	2.70	3.50	5.27	7.90
40	2.16	2.83	4.33	6.41	2.51	3.28	4.91	7.30
50	2.08	2.73	4.18	6.17	2.39	3.15	4.68	6.92
60	2.02	2.66	4.08	6.00	2.31	3.06	4.53	6.66
70	1.97	2.59	3.99	5.85	2.23	2.98	4.41	6.46
80	1.92	2.54	3.92	5.73	2.17	2.91	4.32	6.29
90	1.89	2.50	3.86	5.63	2.12	2.85	4.24	6.15
100	1.85	2.45	3.80	5.54	2.07	2.80	4.17	6.02
∞	1.34	1.84	2.89	4.24	1.34	1.84	2.89	4.24

(Continued)

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution (Continued)

Skew = 0.60

Sample Size	Probability, E = 0.90				Probability, E = 0.95			
	Proportion				Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	1.91	2.61	4.21	6.44	2.07	2.83	4.55	6.92
40	1.83	2.52	4.07	6.23	1.97	2.71	4.35	6.63
50	1.78	2.46	3.98	6.10	1.91	2.62	4.22	6.43
60	1.75	2.42	3.91	6.00	1.86	2.56	4.12	6.30
70	1.72	2.38	3.85	5.91	1.82	2.51	4.05	6.19
80	1.69	2.35	3.80	5.84	1.79	2.47	3.99	6.09
90	1.67	2.32	3.76	6.78	1.76	2.43	3.92	6.01
100	1.65	2.29	3.72	5.72	1.74	2.40	3.88	5.94
∞	1.34	1.91	3.15	4.82	1.34	1.91	3.15	4.82

(Continued)

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution (Continued)

Skew = 0.60

Sample Size	Probability, E = 0.99				Probability, E = 0.999			
	Proportion				Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	2.38	3.21	5.11	7.92	2.81	3.75	5.86	9.14
40	2.24	3.04	4.84	7.45	2.62	3.51	5.50	8.45
50	2.16	2.94	4.68	7.16	2.50	3.36	5.27	8.02
60	2.10	2.86	4.56	6.95	2.41	3.26	5.11	7.72
70	2.04	2.79	4.45	6.78	2.33	3.17	4.97	7.49
80	2.00	2.73	4.37	6.64	2.27	3.10	4.87	7.29
90	1.95	2.68	4.30	6.52	2.21	3.03	4.76	7.12
100	1.92	2.63	4.23	6.40	2.16	2.98	4.68	6.99
∞	1.34	1.91	3.15	4.82	1.34	1.91	3.15	4.82

(Continued)

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution (Continued)

Skew = 0.80

Sample Size	Probability, E = 0.90				Probability, E = 0.95			
	Proportion				Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	1.93	2.77	4.65	7.28	2.12	3.01	5.03	7.88
40	1.86	2.67	4.48	7.06	2.02	2.88	4.80	7.53
50	1.82	2.60	4.38	6.91	1.96	2.79	4.65	7.31
60	1.78	2.56	4.30	6.80	1.91	2.72	4.55	7.15
70	1.75	2.52	4.23	6.70	1.87	2.67	4.46	7.02
80	1.73	2.48	4.18	6.61	1.83	2.62	4.38	6.90
90	1.70	2.45	4.13	6.54	1.80	2.58	4.32	6.80
100	1.68	2.42	4.08	6.46	1.77	2.54	4.26	6.70
∞	1.33	1.96	3.39	5.37	1.33	1.96	3.39	5.37

(Continued)

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution (Continued)

Skew = 0.80

Sample Size	Probability, E = 0.99				Probability, E = 0.999			
	Proportion				Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	2.48	3.44	5.68	9.02	2.91	3.98	6.44	10.46
40	2.32	3.26	5.37	8.49	2.72	3.72	6.05	9.65
50	2.23	3.13	5.18	8.15	2.61	3.56	5.80	9.15
60	2.15	3.05	5.04	7.90	2.52	3.44	5.63	8.80
70	2.08	2.97	4.93	7.70	2.43	3.35	5.50	8.53
80	2.04	2.91	4.83	7.54	2.38	3.26	5.38	8.30
90	2.00	2.85	4.75	7.40	2.32	3.19	5.28	8.12
100	1.96	2.80	4.68	7.28	2.27	3.13	5.19	7.95
∞	1.33	1.96	3.39	5.37	1.33	1.96	3.39	5.37

(Continued)

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution (Continued)

Skew = 1.00

Sample Size	Probability, E = 0.90 Proportion				Probability, E = 0.95 Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	2.02	2.91	5.07	8.11	2.20	3.16	5.51	8.79
40	1.92	2.82	4.89	7.86	2.09	3.03	5.25	8.41
50	1.86	2.75	4.77	7.70	2.02	2.94	5.08	8.18
60	1.82	2.70	4.68	7.58	1.96	2.87	4.97	8.00
70	1.78	2.65	4.61	7.47	1.92	2.82	4.87	7.85
80	1.75	2.61	4.54	7.37	1.87	2.76	4.78	7.72
90	1.72	2.57	4.48	7.28	1.84	2.72	4.71	7.60
100	1.69	2.54	4.43	7.20	1.81	2.68	4.65	7.50
∞	1.30	2.00	3.60	5.91	1.30	2.00	3.60	5.91

(Continued)

Table 10. Empirical Tolerance Factors for a Pearson Type III Distribution (Continued)

Skew = 1.00

Sample Size	Probability, E = 0.99				Probability, E = 0.999			
	Proportion				Proportion			
	0.90	0.95	0.99	0.999	0.90	0.95	0.99	0.999
30	2.59	3.65	6.24	10.15	3.00	4.22	7.00	11.70
40	2.41	3.44	5.89	9.51	2.79	3.93	6.60	10.80
50	2.30	3.32	5.67	9.13	2.66	3.75	6.36	10.23
60	2.22	3.22	5.52	8.85	2.56	3.63	6.18	9.84
70	2.15	3.14	5.40	8.62	2.48	3.53	6.04	9.53
80	2.09	3.08	5.29	8.44	2.41	3.44	5.92	9.28
90	2.04	3.02	5.20	8.28	2.35	3.37	5.82	9.06
100	1.99	2.96	5.12	8.14	2.30	3.30	5.73	8.87
∞	1.30	2.00	3.60	5.91	1.30	2.00	3.60	5.91

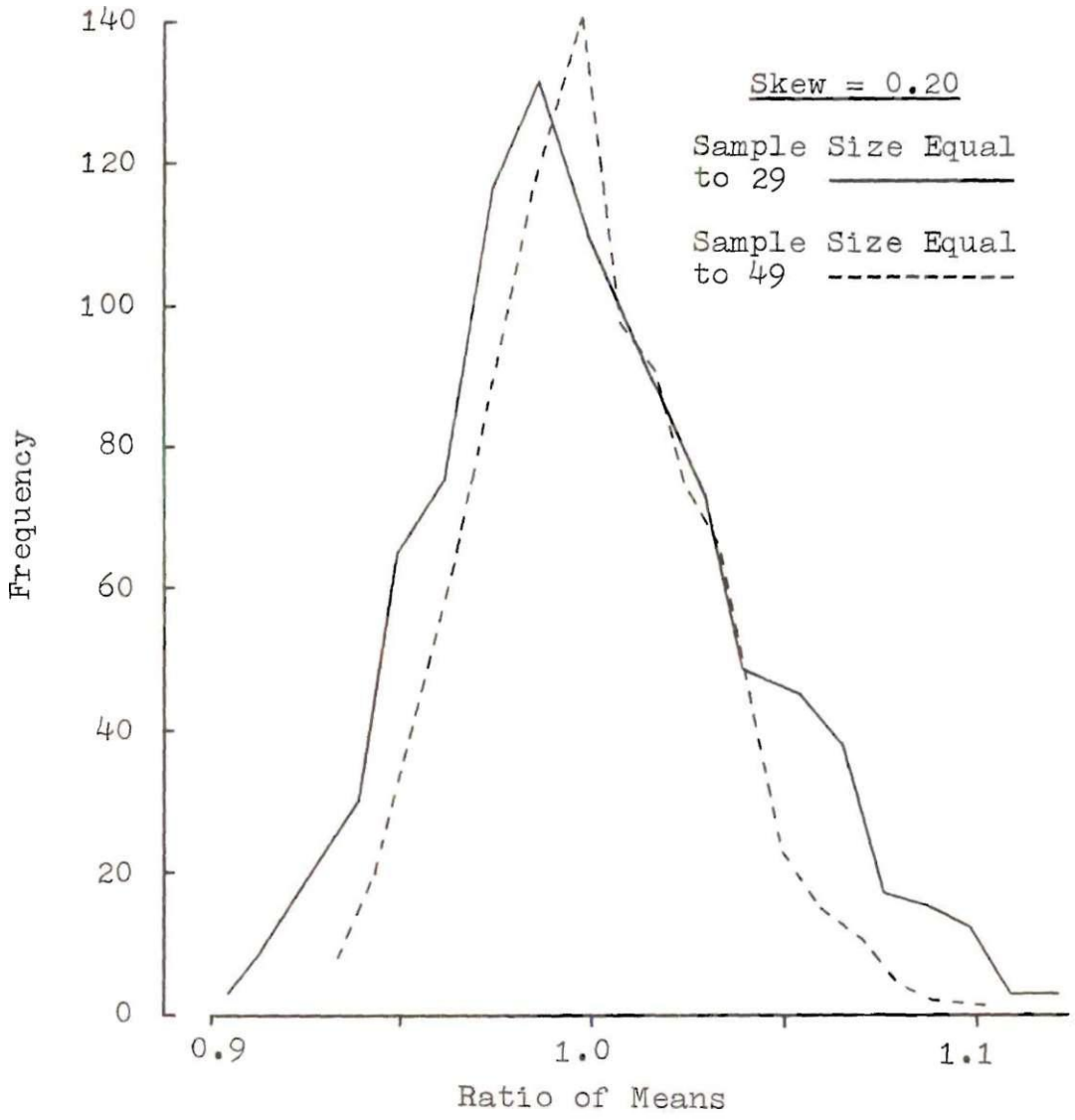


Figure 11. Examples of the Distribution of the Ratio of Means.

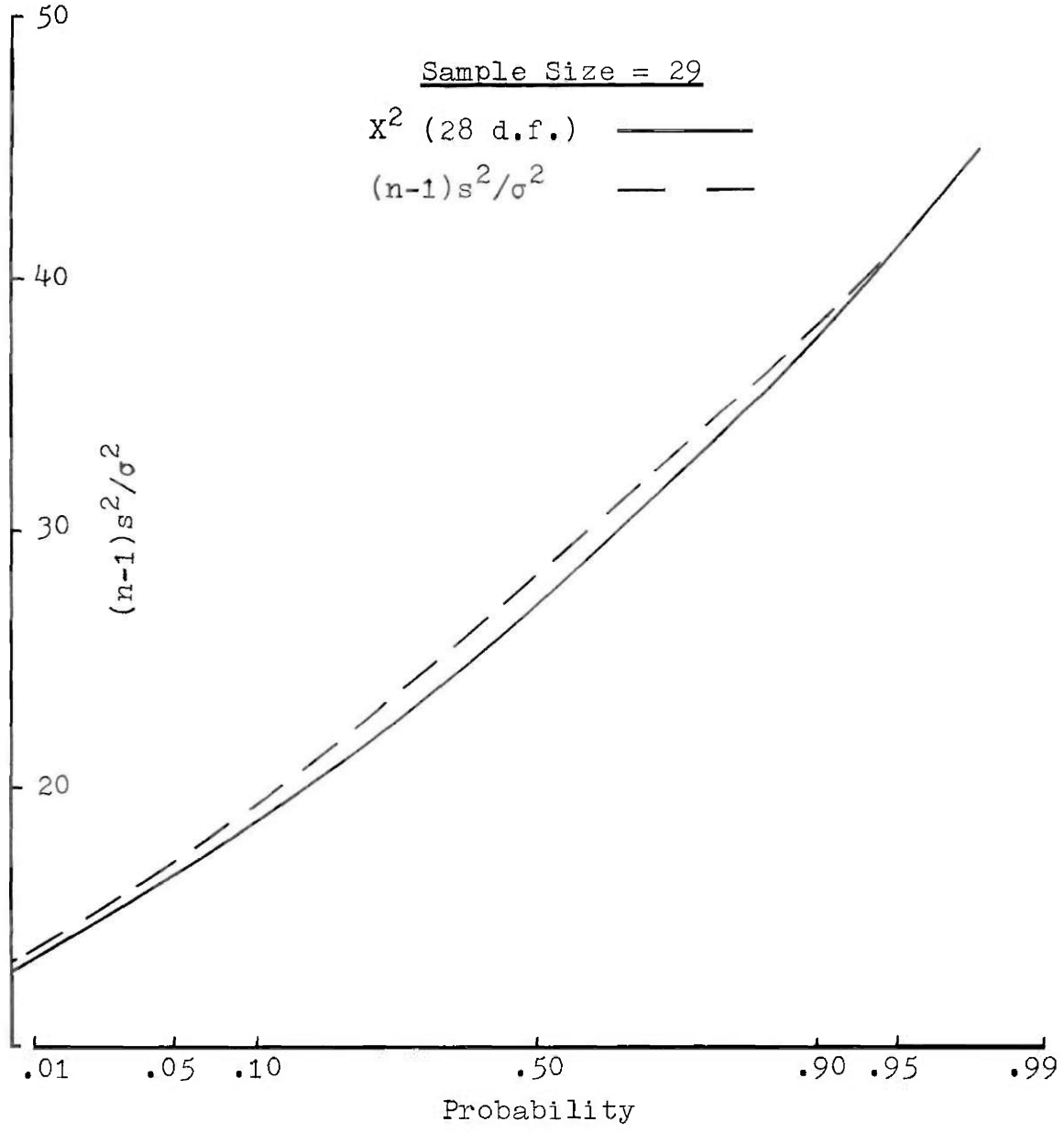


Figure 12. Example of the Percentage Points of $(n-1)s^2/\sigma^2$ and the Theoretical X^2 Values.

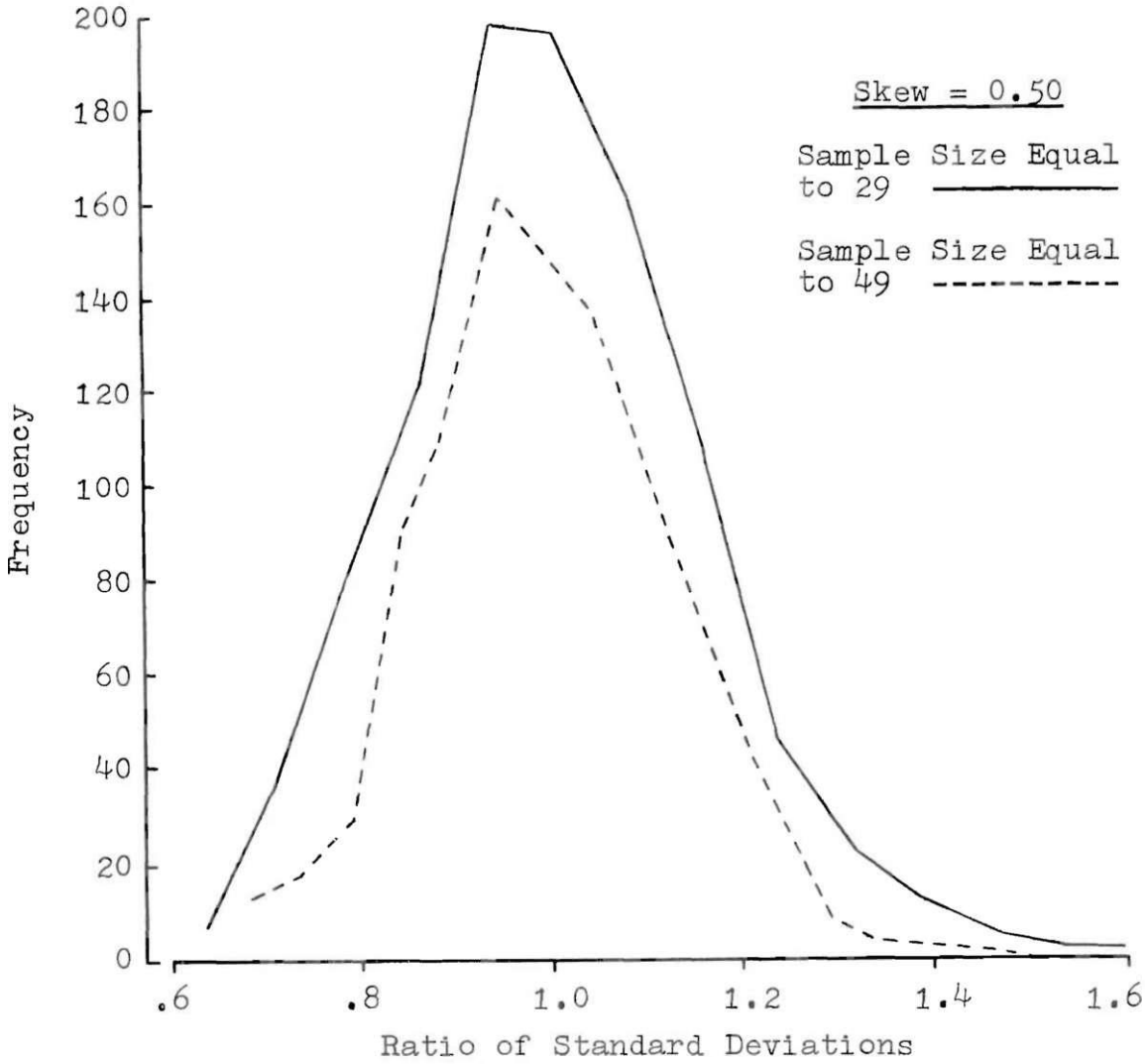


Figure 13. Examples of the Distribution of the Ratio of Standard Deviations.

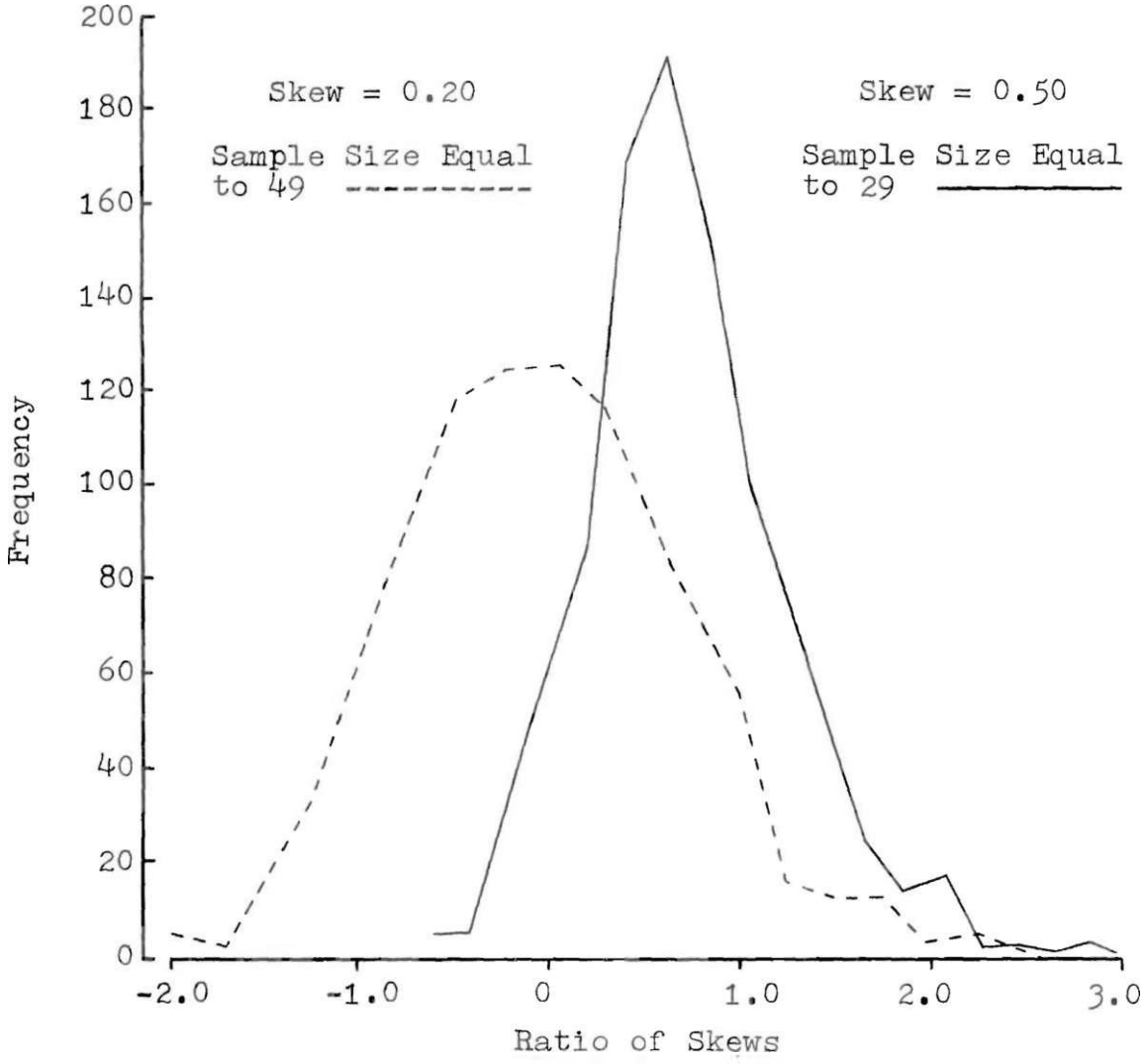


Figure 14. Examples of the Distribution of the Ratio of Skews.

APPENDIX E

REFERENCES

"A Uniform Technique for Determining Flood Flow Frequencies," Bulletin No. 15, Water Resources Council, Washington, D. C., December, 1967.

Bowker, A. H. and Lieberman, G. J., Engineering Statistics, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1959.

Burington, R. S., and May, D. C., Handbook of Probability and Statistics With Tables, Handbook Publishers, Inc., Sandusky, Ohio, 1958.

Elderton, W. P. and Johnson, N. L., Systems of Frequency Curves, Cambridge University Press, London, 1969.

Finney, D. J., Statistics for Mathematicians, Oliver and Boyd, London, 1968.

Fisher, R. A., "The Moments of the Distribution for Normal Samples of Measures of Departure from Normality," Proceedings of the Royal Society of London (A), Vol. 130, 1931, pp. 16-28.

Foster, H. A., "Theoretical Frequency Curves," Transactions of the American Society of Civil Engineers, Vol. 87, 1924, pp. 142-203.

Guttman, I., and Wilks, S. S., Introductory Engineering Statistics, John Wiley and Sons, Inc., New York, 1965, p. 144.

Harter, H. L., "A New Table of Percentage Points of the Pearson Type III Distribution," Technometrics, Vol. 11, No. 1, Feb., 1969, pp. 177-187.

Johnson, N. L., and Welch, B. L., "Applications of the Non-Central t-Distribution," Biometrika, Vol. 31, 1940, pp. 362-376.

Kendall, M. G., and Stuart, A., The Advanced Theory of Statistics, 2nd ed., Vol. 1, Hafner Publishing Co., New York, 1963, p. 62.

Linsley, R. K., Kohler, M. A., and Paulhus, J. L., Hydrology For Engineers, McGraw-Hill Book Company, Inc., New York, 1958.

Markovic, R. D., "Probability Functions of Best Fit to Distributions of Annual Precipitation and Runoff," Hydrology Papers No. 8, Colorado State University, Fort Collins, Colorado, August, 1965.

Matalas, N. C., and Benson, M. A., "Note on the Standard Error of the Coefficient of Skewness," Water Resources Research, Vol. 4, No. 1, Feb., 1968, pp. 204-205.

"Methods of Flow Frequency Analysis," Bulletin No. 13, Inter-Agency Committee on Water Resources, Washington, D. C., April, 1966.

Natrella, M. G., Experimental Statistics, National Bureau of Standards Handbook 91, Washington, D. C., 1963.

Naylor, T. H., et al., Computer Simulation Techniques, John Wiley and Sons, Inc., New York, 1966.

Pearson, E. S., "Some Problems Arising in Approximating to Probability Distributions Using Moments," Biometrika, Vol. 50, 1963, pp. 95-112.

Pearson, K., ed., Tables of The Incomplete Gamma Function, Cambridge University Press, London, 1946.

Selby, S. M., ed., Standard Mathematical Tables, 14th ed., The Chemical Rubber Co., Cleveland, Ohio, 1965.

Snyder, W. M., "Elements of Hydrology for a Program of Flood Insurance," a report prepared for the Department of Housing and Urban Development, Office of Program Policy, Studies of Natural Disasters, July, 1966.

Wald, A., and Wolfowitz, J., "Tolerance Limits for a Normal Distribution," Annals of Mathematical Statistics, Vol. 17, 1946, pp. 208-215.